scientific reports

OPEN

Decision theory for precision therapy of breast cancer

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

Correctly estimating the hormone receptor status for estrogen (ER) and progesterone (PGR) is crucial for precision therapy of breast cancer. It is known that conventional diagnostics (immunohistochemistry, IHC) yields a significant rate of wrongly diagnosed receptor status. Here we demonstrate how Dempster Shafer decision Theory (DST) enhances diagnostic precision by adding information from gene expression. We downloaded data of 3753 breast cancer patients from Gene Expression Omnibus. Information from IHC and gene expression was fused according to DST, and the clinical criterion for receptor positivity was re-modelled along DST. Receptor status predicted according to DST was compared with conventional assessment via IHC and gene-expression, and deviations were flagged as questionable. The survival of questionable cases turned out significantly worse (Kaplan Meier $p < 1\%$) than for patients with receptor status confirmed by DST, indicating a substantial enhancement of diagnostic precision via DST. This study is not only relevant for precision medicine but also paves the way for introducing decision theory into OMICS data science.

Background and significance. Precision medicine relies on biomarkers for selecting the most adequate therapy for each single patient[1–3]. In particular, individually optimized treatment of breast cancer patients has widely been studied, drawing on molecular subtyping[4–7], networks and pathways[8–12] or more general gene expression signatures[13–20]. Among such signatures, PAM50[21,22], PREDICT[23], the Gene expression Grade Index[24] and the Genomic Grade Index[25] have been widely recognized. For a survey see Huang, C. C. et al.[26].

Hormone receptor status (estrogen, ER and progesterone, PGR) as well as the human epidermal growth factor 2 receptor (HER2) are considered most important[27,28] and are estimated by immunohistochemistry (IHC) in clinical practice[29–36]. If at least one of both receptors (ER or PGR) is found positive (i.e. receptors are present), hormone therapy will suffice for an effective treatment, avoiding the much more severe side effects of chemotherapy. However, if the receptor status should accidentally be estimated falsely positive, hormone treatment will not work and the patient may be deprived from life-saving chemotherapy.

Due to its focal importance, receptor assessment quality has been scrutinized in multicenter studies[37,38], suggesting up to 20% misclassifications[39–41]. Guidelines have been issued[39,42,43] but improving assessment precision remains a key issue[44].

A most natural way to increase reliability is to enhance IHC-estimates by additional information, e.g. from gene expression. Methodological advances were reported for data normalization[45–48] as well as for gene expression analysis[49–51]. A survey was given by Bolstad, Irizarry, Astrand and Speed[52]. Chen modelled the probability distributions for gene-expressions of ER, PGR and HER2 and set cutoff values at a posterior probability of 0.5 to discriminate receptor positive from negative patients[53]. Likewise, Laas set cutoffs for receptor positivity within the frequency distributions of gene-expression of ER and HER2[41]. Gong used random sampling to estimate optimum thresholds for ER and HER2 and verified these in a test cohort[54]. Bergqvist set thresholds for HER2 expression according to visual inspection[55]. Lopez introduced fuzzy rules to identify breast cancer biomarkers[56]. Concluding, prediction algorithms for low dimensional input space (such as hormone receptors), yield fairly similar results.

In previous papers[57,58] we have worked on such approaches, applying standard statistical means (odds-products). In this work we expand the approach by drawing on Dempster Shafer decision Theory (DST)[59].

DST has been widely applied in self driving cars[60–62], driver's vigilance monitoring[61], aircraft technology[63,64] and also in medical settings, e.g. image based decisions[65], diagnosis of prostate[66] and breast cancer[67]. In all these

[1]Section of Biosimulation and Bioinformatics, Center for Medical Statistics, Informatics and Intelligent Systems (CeMSIIS), Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria. [2]Translational Gynecology Group, Department of Obstetrics and Gynecology, Comprehensive Cancer Center, Medical University of Vienna, Waehringer Guertel 18-20, 1090 Vienna, Austria. [3]Department of General Gynecology and Gynecologic Oncology, Medical University of Vienna, Waehringer Guertel 18-20, 1090 Vienna, Austria. ✉email: wolfgang.schreiner@meduniwien.ac.at

			Logistic regression parameters		Logistic regression quality		Upper limits for beliefs	
	Probe set	c_0	c_1	Deviance of fit	N of samples	$\hat{\alpha}$	$\hat{\beta}$
Estrogen								
Gene	ESR1	205225_at	3.798	−2.302	1106.4	2227	0.820	0.885
Co-gene	AGR3	228241_at	1.528	−1.557	1254.1		0.790	0.841
Progesterone								
Gene	PGR	208305_at	0.886	−1.944	1329.3	1714	0.755	0.660
Co-gene	ESR1	205225_at	4.112	−2.422	845.0		0.826	0.878

Table 1. Receptor genes, co-genes and parameters from logistic regression. Probe sets refer to the Affymetrix chip U133A + 2.0. For 'deviance of fit', see p.118 in McCullagh[81]. Note the double role of ESR1: It is the very receptor gene for ER but also serves as best co-gene for PGR, see our previous work[57] regarding the selection of co-genes. Upper limits for beliefs ($\hat{\alpha}$, $\hat{\beta}$) are explained in the next section (Eqs. 4–6).

applications, unclear or even contradicting information from several sources are fused to arrive at decisions of optimized precision.

Using decision theory for receptor status assessment. Classical probability theory characterizes the chance for an event to occur by a single number, its probability. A probability, p, characterizes the event as such, including its circumstances, but disregards the measuring process itself. The performance of the measuring process is characterized by sensitivity and specificity, in short by its receiver operating characteristics (ROC). In contrast, DST includes the measuring process and characterizes expected outcomes (say 'receptor positive', 'Rez⁺') by two numbers, the belief bel(Rez⁺) and the plausibility, pl(Rez⁺), together called 'evidence' of the outcome. Note that this definition of evidence specifically relates to the framework of decision theory and differs from more general meanings of the term, e.g. in 'evidence based medicine'. The belief indicates the chance to correctly measure 'positive' by virtue and quality of the measuring process. The plausibility indicates the chance that the reading 'positive' (for an actually positive receptor) could also result (a) by chance or (b) represent a false positive outcome (for a receptor status truly 'negative'). For each measurement with dichotomic outcome $\Omega = \{\text{Rez}^+, \text{Rez}^-\}$, $\text{pl}(\text{Rez}^+) = 1 - \text{bel}(\text{Rez}^-)$. DST additionally accommodates uncertainty, θ, defined as $\theta(\text{Rez}^+) = \text{pl}(\text{Rez}^+) - \text{bel}(\text{Rez}^+) = 1 - \text{bel}(\text{Rez}^-) - \text{bel}(\text{Rez}^+)$, i.e. the uncertainty of an event is the difference between its plausibility and its belief. For brief notation we use $\text{bel}(\text{Rez}^+) = \alpha$ and $\text{bel}(\text{Rez}^-) = \beta$.

The main parts of this paper address readers familiar with DST and demonstrate how to combine several evidences for single hormone receptors and further combine several receptor estimates towards an overall hormone receptor status. Readers not yet familiar with DST may first see the 'supplementary methods' and then resume to read the following chapters.

Materials and methods
Data curated and used.
In order to establish a comprehensive database[68], we decided to re-use published gene expression data[69] and screened the Gene Expression Omnibus (GEO)[70,71] for breast cancer studies using the Affymetrix chip U133A + 2.0[72]. We found and curated 38 studies with 3753 samples.

Out of numerous methods available for normalization[46,73–76] as compared by Bolstad[52] and evaluated in our previous work[77], we used RMA (MATLAB affyrma)[78] and standardized data for each sample. Further batch corrections have been scrutinized in our previous work[77] and were found ambiguous. We therefore refrained from performing them.

We adopted receptor genes and selected co-genes (see Table 1) from our previous paper[57]. Whereas in multiple logistic regression all genes would be incorporated simultaneously in the predictor (multiple logistic regression), the ODDS method performs logistic regressions on single genes and then combines the odds. In other words, in this former work, probabilities from IHC, gene and co-gene were joined according to conventional statistics, i.e. by multiplying odds. Hence we refer to this method as 'ODDS' in the following. Like most binary classifiers[79], ODDS builds a score out of input variables (i.e. gene expression). This score is either compared to a threshold and the corresponding class assigned (crisp classification, + or −). Or else, the sharp threshold is replaced by an interval, within which the score is considered 'uncertain'. We chose the latter possibility in the setup of ODDS and excluded samples classified 'uncertain'. This lets the remaining samples (+, −) gain certainty in their assessment, which is appropriate for a data basis on which our new discrimination method (DST) is evaluated.

When performing ODDS like in our previous paper, results may contradict IHC—for a few samples. As opposed to this, in the current work we took IHC-estimates for granted and applied ODDS only to samples where no IHC estimate was available. Hence, values imputed by ODDS could never contradict IHC-estimates—just to be on the save side. Moreover, 'uncertain' results from ODDS were also discarded. All in all, to establish the database for the present work, we retained results either from IHC or safely imputed by ODDS and call this procedure 'sODDS' (*safe*Odds).

As first step of data cleansing we ruled out crosstalk by HER2: We dismissed HER2⁺ samples and accepted only those definitely assessed negative, either by IHC (1798 samples) or safely imputed via ODDS (1010 samples). For
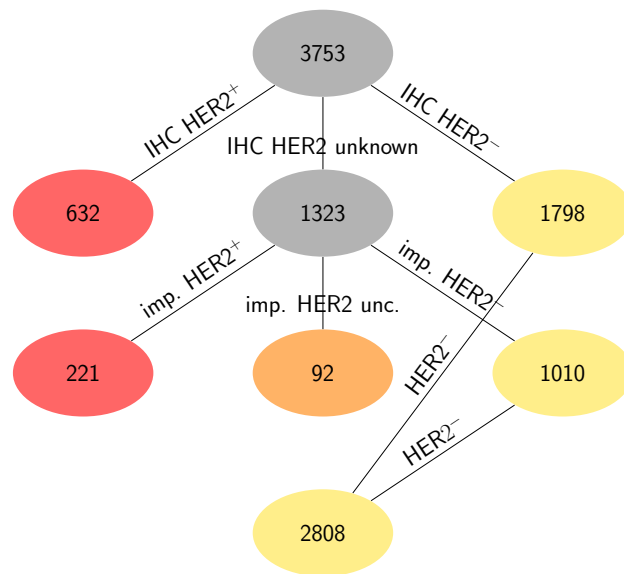
Figure 1. Selection of HER2-negative (HER2⁻) samples. The sODDS-method[57] was applied to 3753 patient samples to extract those HER2⁻. If IHC receptor status was available, it was taken as is and samples accepted (1798 HER2⁻, yellow) or excluded (632 HER2⁺, red). Patients with unknown IHC-status (1323, grey) were accepted if HER2⁻ could be safely imputed (imp. HER2⁻) by the ODDS method (1010, yellow) and excluded if ODDS safely yielded HER2⁺ (221, red) or 'uncertain' (imp. HER2 unc., 92, beige). All in all, 1798 + 1010 = 2808 HER2⁻ patients resulted.

imputation we used default parameters and the threshold $|score| > \log((1 - \varepsilon)/\varepsilon)$ with $\varepsilon = 0.15$, like in our previous work[57]. Finally, 1798 + 1010 = 2808 samples remained as HER2⁻, according to the sODDS method, see Fig. 1.

These 2808 samples were then checked regarding ER and PGR as follows: If IHC-estimates were available for ER as well as PGR, samples were taken as is (1714 samples). If IHC was only available for ER and the status of PGR could be safely imputed via ODDS[57] with sufficient score (see above), samples were also included (513 samples). If IHC was missing for both, ER as well as PGR, samples were taken only if both status could be imputed safely via ODDS (332 samples). Thus, 1714 + 513 + 332 = 2559 samples were retained with receptor status established most accurately by state of the art (IHC plus ODDS ≙ sODDS), see Fig. 2. We used only these to evaluate the benefit of applying decision theory.

Survival rates and quality of previous receptor status assessments.   A study-wise evaluation of receptor status assessment via IHC and ODDS was performed, see Supp. Table 2. The concordance of receptor status between IHC and ODDS is much better for ER than for PGR, due to larger variance of ER in gene expression data. Results were further elaborated in a control chart, see Fig. 3. The dotted line represents p, the overall average rate of differences, computed from Supp. Table 2 as $p = 159 / 2227 \sim 0.071$. Markers within bars represent expectation values $(\Delta_{\text{IHC, ODDS}} + 1)/(N_{\text{IHC}} + 2)$ of discordance rates derived from the beta-distribution, see the mathematical formalism in section 'Concordance between IHC and gene expression' in supplementary materials. Bars denote 95% confidence intervals of the respective mean values, allowing the qualification of studies against each other: If an upper bound lies below the dotted line in Fig. 3, the respective study has a discrepancy rate significantly below average (green). If a lower bound lies above the dotted line, the discrepancy rate of that study is significantly above average (red).

Similarly, a control chart for progesterone is shown in Supp. Fig. 2.

Additionally, we evaluated the scmgene-marker (from package genefu[80]) for estrogen only, since scmgene does not give PGR estimates. Note that scmgene was evaluated for all (3753) samples but results were only compared for those 2559 with HER2⁻ and reliable hormone receptor status, see Supp. Table 1.

Survival has been related to receptor status for ER (Supp. Fig. 3), PGR (Supp. Fig. 4) and hormone overall, i.e. either estrogen or progesterone being positive, see Fig. 4.

Responsibility functions for gene expression.   In order to compute DST evidences from gene expression, we will first define responsibility functions as a prerequisite. The concept was coined by Hastie[79] for weighing two distributions against each other regarding membership of a measurement in question. We use this for distributions of gene-expression data and their indication of receptor status. Based on these, 'mass functions' will be computed, and these converted into evidence:

Gene expression + IHC → responsibility functions (→ mass functions) → evidence.

In the main part of this paper for brevity we skip the concept of masses (hence above shown in parenthesis) and present formulae for evidence directly based on responsibility functions. Masses are relegated to supplementary methods.
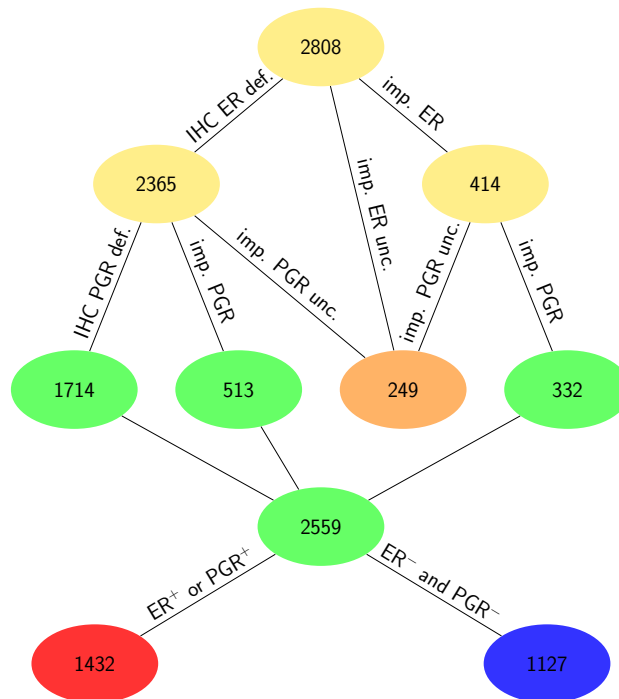
Figure 2. Selection of samples with definite hormone receptor status. Selection started from 2802 samples HER2⁻, see Fig. 1. If an IHC estimate for estrogen (ER) was available, ER was considered defined (ER def., 2365). Otherwise the ODDS method[57] was applied, yielding either a safe result (imp. ER, 414) or resulting uncertain (imp. ER unc.), contributing to the set of 249 samples shown in beige. ODDS was applied to the set of 414 samples to impute progesterone status, either safely (imp. PGR, 332, green) or ending up uncertain (imp. PGR unc.), contributing to the beige set of 249. Finally, out of the 2365 samples with IHC available, IHC for progesterone was either available (IHC PGR def., 1714, green), safely imputed by ODDS (imp. PGR, 513, green) or uncertain (imp. PGR unc.), again contributing to the beige set. All in all, IHC receptor status was available for 1714 samples and safely imputed for $513 + 332 = 845$ samples. This resulted in 2559 samples with definite hormone receptor status, out of which 1432 were positive (red) and 1127 negative (blue). These constituted the input set for DST, as shown in Fig. 7. Note that no samples exist with PGR_{IHC} defined and ER_{IHC} imputed.

Responsibility functions are exemplified now for the expression of *one* gene, labelled 'Expr', to keep notation slim. It applies analogously to the other genes considered.

We construct the responsibility function, r_+, for receptor positivity from logistic regression (fitting coefficients: c_0, c_1) of IHC receptor status versus gene expression, x_{Expr}, see Eq. (1) and the increasing dotted red line in Fig. 5. The (dual) responsibility function, r_-, for receptor-negativity is simply given by $r_- = 1 - r_+$, see Eq. (1) and the decreasing blue dotted line in Fig. 5.

$$r_+ \left(x_{Expr} | c_0, c_1 \right) = \frac{\exp \left(c_0 + c_1 x_{Expr} \right)}{1 + \exp \left(c_0 + c_1 x_{Expr} \right)}$$
$$r_- \left(x_{Expr} | c_0, c_1 \right) = 1 - r_+ \left(x_{Expr} | c_0, c_1 \right)$$

(1)

Logistic regression is carried out separately for estrogen receptor gene (ESR1), estrogen co-gene (AGR3), progesterone receptor gene (PGR) and its co-gene (ESR1). Results are consolidated in Table 1.

Obtaining evidences. We define DST masses (see supplementary methods) and evidences by drawing on responsibility functions as follows: We observe in Fig. 5 that $r_+ \rightarrow 1$ as gene expression approaches its maximum. However, not even maximum gene expression in reality provides total certainty of receptor status being positive. To account for this fact, the responsibility function has to be multiplied by an upper bound, $\hat{\alpha}_{Expr}$, for the belief in '+'. Similarly $\hat{\beta}_{Expr}$ acts as upper bound for the belief in '−'. Next we show how $\hat{\alpha}_{Expr}$ and $\hat{\beta}_{Expr}$ are obtained, see Fig. 6.

Out of all positive IHC-measurements (represented by the interval (0,1)), a certain fraction has been confirmed positive by gene expression, due to virtue of the method. This fraction of measurements is represented by the interval $(0, \alpha_{Expr})$ as part of the interval (0,1). Given a positive IHC measurement, the plausibility $\beta_{Expr} = 0$, and hence the rest of the interval (0,1) represents nothing but uncertainty according to DST: $\theta_{Expr} = 1 - \alpha_{Expr} - 0$. As depicted in Fig. 6, some fraction, θ^+, of uncertain predictions from gene expression will result positive merely by chance, not by virtue of the method. Both parts together represent all true positive
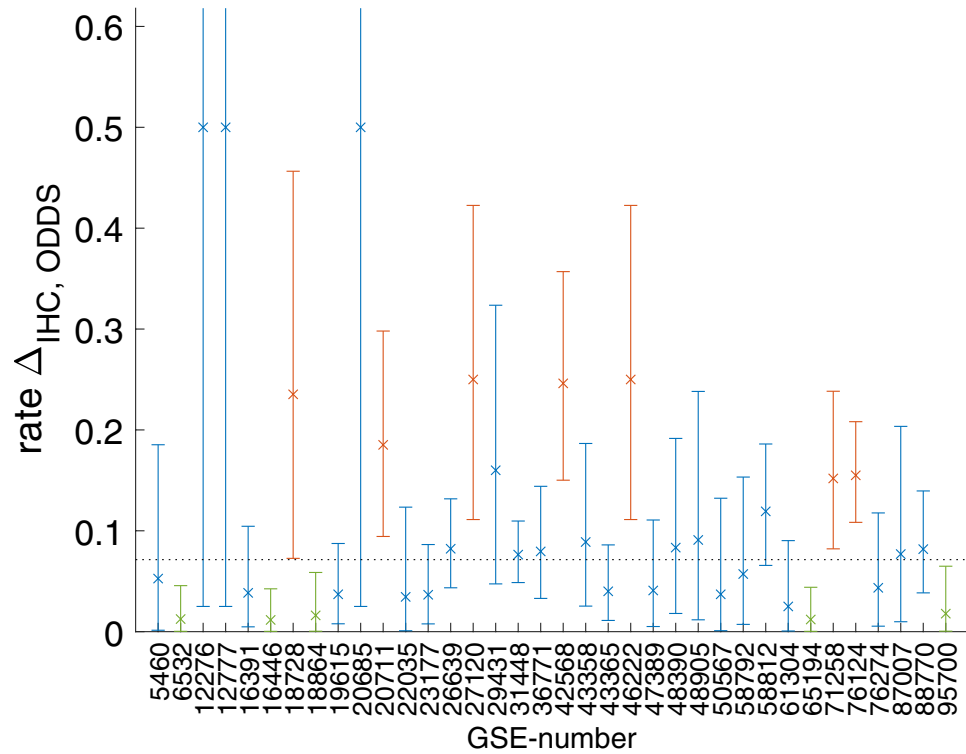
Figure 3. Control chart for estrogen receptor status: differences between IHC and ODDS gene expression estimates. x-axis: label of study. y-axis: rate of difference. Dotted line: overall mean rate of differences between IHC and ODDS $p = 159/2227 = 0.071$, see SuppTable 2. Markers (x) within error bars: rate of discordant samples in respective study. Lower and upper bounds of error bars denote 95% confidence limits. If the upper bound lies below the dotted line, the respective study has a discrepancy rate significantly below average (green). If the lower bound lies above the dotted line, the respective study has a discrepancy rate significantly above average (red).
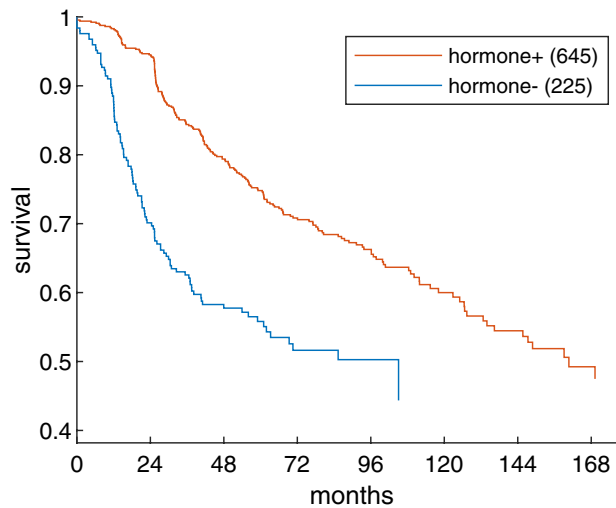

Figure 4. Survival free from recurrence for hormone status. Hormon status is considered positive, if either ER or PGR are positive, otherwise negative. Kaplan Meier estimates of survival free from recurrence for 870 patients (who had rfs-data) for different hormone receptor status. Log-rank test $p = 4.2e–06$.

IHC-measurements: $\alpha_{\text{Expr}} + \theta^{+}_{\text{Expr}} = TP$. A second part of uncertainty, θ^{-}_{Expr}, will represent erroneously assessed false positives (*FP*), although being truly negative samples. Based on this nomenclature we proceed as follows:

Consider all samples qualified 'uncertain' by gene expression, see Fig. 6. The part θ^{+}_{Expr} represents truly positive samples, whereas θ^{-}_{Expr} are truly negative ones. We may now assume on good grounds that the ratio
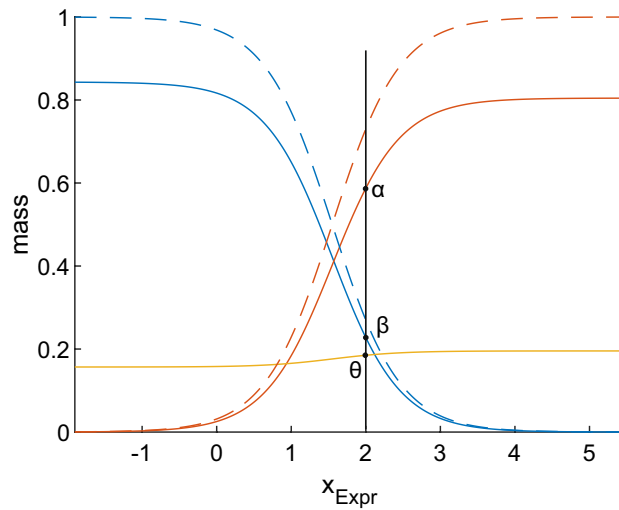
Figure 5. Decision theory evidences obtained from logistic regression. A logistic regression of IHC receptor status (IHC$^+$ $\triangleq$ 1, IHC$^-$ $\triangleq$ 0) versus gene expression (x_{Expr}) was performed to obtain the responsibility function for receptor positivity r_+ (dotted red curve) and r_- (dotted blue). It will be shown later (Eq. 6) that r_+ has to be multiplied by an upper limit, $\widehat{\alpha}_{\text{IHC}}$, to obtain the actual belief $\alpha_{\text{Expr}}(x_{\text{Expr}}) = \widehat{\alpha}_{\text{IHC}} \cdot r_+(x_{\text{Expr}})$, see the solid red curve. Likewise $\beta_{\text{Expr}}(x_{\text{Expr}}) = \widehat{\beta}_{\text{IHC}} \cdot r_-(x_{\text{Expr}})$ (solid blue). Uncertainty: ochre. For a given expression value, e.g. $x_{\text{Expr}} = 2$, one can read off belief in positive (α), belief in negative (β) and uncertainty (θ).
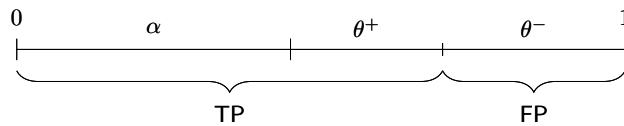
Figure 6. Estimating an upper bound for belief in receptor positivity. True positives (TP) and false positives (FP) can be used to derive a first equation in α_{Expr} and θ_{Expr} (Eq. 3). A second equation may be deduced by including TN and FN, Eq. (2). Both can be solved to estimate α_{Expr} and θ_{Expr}. From these follow upper bounds for the belief in positivity ($\hat{\alpha}_{\text{Expr}}$) and negativity ($\hat{\beta}_{\text{Expr}}$), Eqs. (4 and 5), respectively.

$\theta_{\text{Expr}}^+ \big/ \theta_{\text{Expr}}^-$ among the uncertain ones from gene expression is similar (equal) to the ratio of positives and negatives found by IHC, see also Shoyaib[82]. Considering the contingency table of IHC-measurements, note that positive measurements comprise $TP + FN$, whereas negative ones are composed as $TN + FP$. Hence we may put

$$\frac{\theta_{\text{Expr}}^+}{\theta_{\text{Expr}}^-} = \frac{TP + FN}{TN + FP} \tag{2}$$

Moreover, the fraction of gene expression measurements being positive by virtue or chance, $\alpha_{\text{Expr}} + \theta_{\text{Expr}}^+$, is set equal to the fraction of IHC-true positives:

$$\alpha_{\text{Expr}} + \theta_{\text{Expr}}^+ = \frac{TP}{TP + FP} \tag{3}$$

Since $\alpha_{\text{Expr}} + \theta_{\text{Expr}}^+ + \theta_{\text{Expr}}^- = 1$, Eqs. (2 and 3) can now be solved to yield the maximum possible belief in positive receptor status (derived from the gene in question):

$$\hat{\alpha}_{\text{Expr}} = \frac{TP \cdot TN - FP \cdot FN}{(TP + FP) \cdot (TN + FP)} \tag{4}$$

Likewise we obtain for the maximum belief in negative status:

$$\hat{\beta}_{\text{Expr}} = \frac{TP \cdot TN - FP \cdot FN}{(TN + FN) \cdot (TP + FN)} \tag{5}$$

Considering these maximum beliefs, we finally obtain for the evidence derived from gene expression, for illustration see Fig. 5:

$$\alpha_{\mathrm{Expr}}\left(x_{\mathrm{Expr}}\right) = \hat{\alpha}_{\mathrm{Expr}} \cdot r_+\left(x_{\mathrm{Expr}}|c_0, c_1\right)$$
$$\beta_{\mathrm{Expr}}\left(x_{\mathrm{Expr}}\right) = \hat{\beta}_{\mathrm{Expr}} \cdot r_-\left(x_{\mathrm{Expr}}|c_0, c_1\right) \tag{6}$$

Note that the whole formalism (Eqs. 1–6) is performed separately for each gene (and co-gene), using the data of corresponding IHC measurements. For generality and brevity of above notation, the subscript 'Expr' stands for each of those genes.

Completing the decision theory framework.

With responsibility functions and evidences available we may now assemble the decision framework as follows.

Each hormone receptor (ER, PGR) is assessed via three sources of information, each yielding separate evidence (in our case of dichotomous data: 2 numbers):

(1) IHC → evidence (α_{IHC}, β_{IHC})
(2) gene expression of the receptor gene → evidence (α_{Gen}, β_{Gen})
(3) gene expression of a co-gene → evidence (α_{Co}, β_{Co})

In a first step, evidences for the receptor gene (α_{Gen}, β_{Gen}) and its co-gene (α_{Co}, β_{Co}) are combined by the evidence combination rule (ECR) introduced by Dempster[83] (labelled '$\oplus_{\mathrm{D}}$') to obtain joint evidence (α_{Expr}, β_{Expr}) from gene expression:

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

As explained in supplementary methods, the operation $\oplus_{\mathrm{D}}$ yields in detail:

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

This evidence for gene expression is—in a second step—combined with the evidence from IHC, (α_{IHC}, β_{IHC}), either using rule '$\oplus_{\mathrm{D}}$' once more or, alternatively the ECR defined by Yager[83], labelled '$\oplus_{\mathrm{Y}}$'. Since $\oplus_{\mathrm{Y}}$ more easily accommodates contradicting evidence (from IHC and gene expression), we opt for $\oplus_{\mathrm{Y}}$ and obtain:

$$\alpha_{\mathrm{Rez}} = \alpha_{\mathrm{Expr}}\alpha_{\mathrm{IHC}} + \theta_{\mathrm{Expr}}\alpha_{\mathrm{IHC}} + \alpha_{\mathrm{Expr}}\theta_{\mathrm{IHC}}$$
$$\beta_{\mathrm{Rez}} = \beta_{\mathrm{Expr}}\beta_{\mathrm{IHC}} + \theta_{\mathrm{Expr}}\beta_{\mathrm{IHC}} + \beta_{\mathrm{Expr}}\theta_{\mathrm{IHC}}$$
$$\theta_{\mathrm{Rez}} = \theta_{\mathrm{Expr}}\theta_{\mathrm{IHC}} + \alpha_{\mathrm{Expr}}\beta_{\mathrm{IHC}} + \beta_{\mathrm{Expr}}\alpha_{\mathrm{IHC}} \tag{9}$$

In supplementary methods we explain in detail how general concepts of DST (capable of handling a set of multiple outcomes) boil down to Eqs. (8 and 9) in case of just two outcomes, receptor status $\{'+', '-'\}$.

The above procedure is carried out similarly to obtain the combined evidence for the PGR-status (α_{PGR}, β_{PGR}). Note that notation is turned from general to specific in the following ($\alpha_{\mathrm{Rez}} \to \alpha_{\mathrm{ER}}$ and $\alpha_{\mathrm{Rez}} \to \alpha_{\mathrm{PGR}}$, respectively).

In a last step, the clinical decision for 'hormone versus chemo' is modelled along the lines of DST. Clinically, a patient is considered receptor positive (and will receive hormone therapy) if either ER or PGR (or both) are positive. Classically, this is a crisp, logical decision as stated. DST however, allows more elaborate combination rules to combine evidences for estrogen (α_{ER}, β_{ER}) and progesterone (α_{PGR}, β_{PGR}).

Thus, considering ER as well as PGR, each being assessed by IHC, gene and co-gene, performing some algebra (detailed in supplementary methods) yields the overall evidence for 'hormone receptor status' ($\alpha_{\mathrm{H}} = \mathrm{bel}_{\mathrm{H}}(pos)$, $\beta_{\mathrm{H}} = \mathrm{bel}_{\mathrm{H}}(neg)$):

$$\alpha_{\mathrm{H}} = \alpha_{\mathrm{ESR}} + \alpha_{\mathrm{PGR}} - \alpha_{\mathrm{ESR}}\alpha_{\mathrm{PGR}}$$
$$\beta_{\mathrm{H}} = \beta_{\mathrm{ESR}} \cdot \beta_{\mathrm{PGR}} \tag{10}$$

As always we have to accept an amount of uncertainty given by $\theta_{\mathrm{H}} = 1 - \alpha_{\mathrm{H}} - \beta_{\mathrm{H}}$, indicating hormone receptor status being indeterminable. From these evidences, a most reasonable decision rule can be derived:

$$\text{Hormone receptor positive} \leftrightarrow \alpha_{\mathrm{H}} > \beta_{\mathrm{H}} + \theta_{\mathrm{H}}$$
$$\text{Hormone receptor negative} \leftrightarrow \beta_{\mathrm{H}} > \alpha_{\mathrm{H}} + \theta_{\mathrm{H}}$$
$$\text{Hormone receptor indeterminable} \leftrightarrow |\alpha_{\mathrm{H}} - \beta_{\mathrm{H}}| \leq \theta_{\mathrm{H}} \tag{11}$$

In our simple case with only two outcomes these criteria reduce to $\alpha_{\mathrm{H}} > 0.5$ for definitely receptor positive and $\beta_{\mathrm{H}} > 0.5$ for definitely receptor negative.

		Method DST			
		−	Uncertain	+	Σ
Method sODDS	−	1023	103	1	1127
	+	0	50	1382	1432
Σ		1023	153	1383	2559

Table 2. Receptor status due to decision theory (DST) compared with previous method (ODDS). 2559 Patients have been assessed for hormone receptor status by method sODDS to comprise the input dataset. DST was then applied to this dataset. A sample was considered H⁺ if at least one receptor was positive. Difference between methods sODDS and DST is reflected in Cohen's $\kappa = 0.88$. Note that uncertain results were considered as separate category in computing κ.
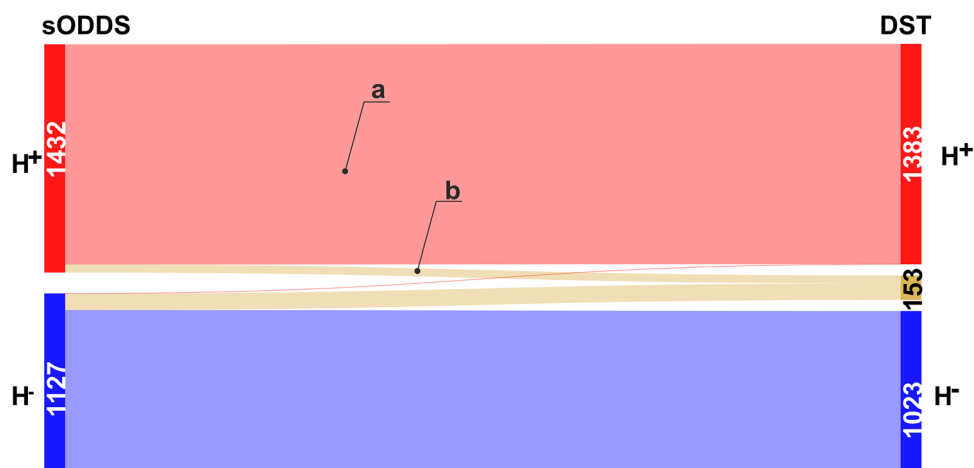
Figure 7. Alluvial flow diagram for patients originally diagnosed by ODDS and then checked by DST. Left margin: Hormone receptor assessed by method sODDS: positive (H⁺) and negative (H⁻). A sample was considered H⁺ if at least one receptor was positive. Lacking IHC estimates were imputed by ODDS and included only if imputation was safe, otherwise discarded. Hence, no patients with receptor status 'undetermined' appear in dataset subjected to DST. Right margin: Evaluation by DST confirms the vast majority of estimates from sODDS but leaves some 153 as undetermined (ochre colored). Only one patient is definitely considered false negative and changed to positive.

Results

DST results for patient cohort. In our previous work[57], we considered receptors (ER, PGR) separately. For each of them routine estimates (based on IHC only, 1714 samples, see Fig. 2) could be significantly improved by adding information from gene expression via conventional statistics (odds ratio) and we label this former method 'ODDS'. Note that for patients lacking one or both IHC estimates, receptor status was imputed from gene expression according to ODDS, including an uncertainty region, as in our previous paper. However, to obtain a dataset of optimum quality for the present work, only samples rendering safe imputations were retained (513 + 332 = 845), see Fig. 2.

Then we demonstrated that further improvement in receptor assessment can be achieved by applying decision theory. Besides yielding new estimates for each receptor, DST allows to combine estimates for ER and PGR into a joined 'hormone' receptor estimate. Based on Eqs. (10 and 11), hormone receptor status was newly predicted by DST for each patient, see the contingency Table 2 and Fig. 7.

To demonstrate the clinical benefit of DST, we consider two groups:

(1) Patients diagnosed receptor positive by sODDS, and being confirmed by DST. They have correctly received hormone treatment, see the flow labelled 'a' in Fig. 7.

(2) Patients diagnosed receptor positive by sODDS, but being questioned by DST, see flow 'b'. They have very probably also received hormone treatment but it remained ineffective. At the same time they might have been deprived of life-saving chemo.

Figure 8 shows a strikingly worse survival of patients assumed receptor positive although being negative (logrank–Wilcoxon $p = 0.009$). Note that out of 1432 patients considered positive by sODDS, only for 651 survival data were available. Among these, 26 were questioned by DST. Nonetheless the difference was statistically significant at the 1%-level.
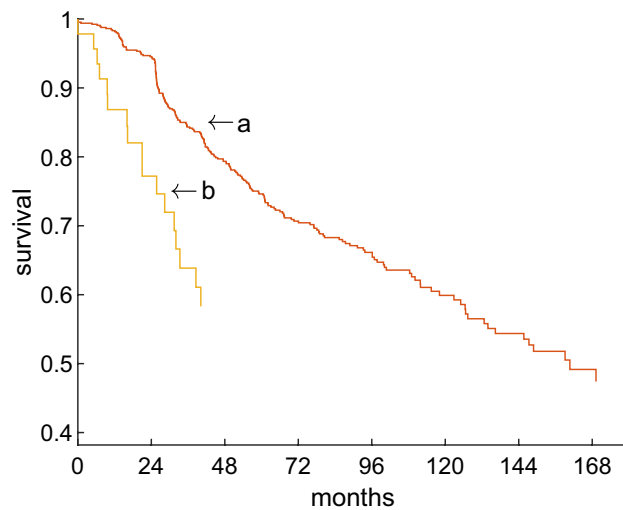
Figure 8. Precise versus inadequate treatment based on hormone receptor status. Survival free from recurrence, according to Kaplan–Meier. Red: Patients treated correctly as receptor positive (flow (a) in Fig. 7). Ochre: Patients erroneously assigned positive (flow (b) in Fig. 7).

		Method ODDS			
		−	Uncertain	+	Σ
Method sODDS	−	1035	52	40	1127
	+	31	30	1371	1432
Σ		1066	82	1411	2559

Table 3. Receptor status according to ODDS. 2559 Patients have been assessed for hormone receptor status by sODDS to comprise the input dataset. A sample was considered H$^+$ if at least one receptor was positive. Difference in methods is reflected in Cohen's $\kappa = 0.88$.

		Method DST			
		−	uncertain	+	Σ
Method ODDS	−	1022	44	0	1066
	Uncertain	1	70	11	**82**
	+	0	39	1372	1411
Σ		1023	**153**	1383	2559

Table 4. DST Decision theory compared to ODDS. The same dataset (2559 samples) was subjected to DST and ODDS, both including the same uncertainty regions as described. DST is seen to flag 153 patients as uncertain, i.e. of questionable hormone receptor status. As compared to ODDS almost double as many patients are detected to be re-checked before therapies are assigned. Difference in methods is reflected in Cohen's $\kappa = 0.88$.

DST compared to ODDS. It is interesting to see what DST achieves as compared to ODDS when starting from the same patient cohort. We therefore took the same sample of patients (2559) as in Table 2 and Fig. 7, with 1432 receptor status positive and 1127 negative according to sODDS. Instead of DST we this time applied the ODDS method, including a uncertainty region, see Table 3.

Most strikingly, DST had flagged almost double as many patients 'uncertain' (156, see Table 2) as compared to ODDS (82), although the very same safety threshold was used. This clearly reflects an advantage of DST, since patients with questionable status would be re-evaluated before assigning the adequate therapy.

To characterize these differences in judgement we performed a face-to-face comparison between DST and ODDS on the very same input dataset, see Table 4. Samples additionally labelled 'uncertain' by DST originate to almost equal parts from negative (44) and positive (39) estimates according to ODDS. Only very few samples considered uncertain by ODDS, migrate to negative (1) or positive (11) according to DST.
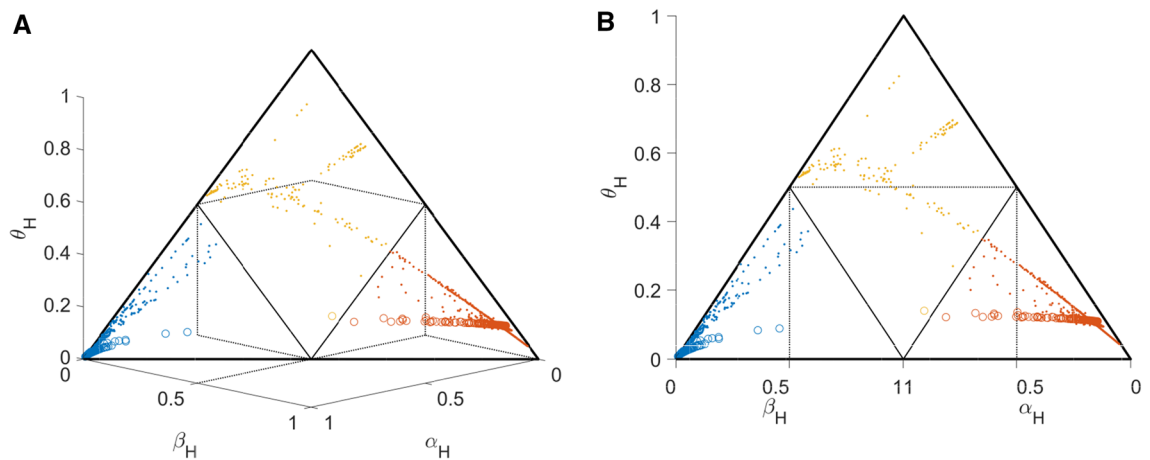
Figure 9. Decision space of hormone receptor status according to DST. (**A**) The three evidences of receptor status, belief in positive (α_H), belief in negative (β_H) and uncertainty (θ_H), of all samples shown in a three dimensional plot. Colors indicate receptor assignment according to DST (positive $\triangleq$ red, negative $\triangleq$ blue, uncertain $\triangleq$ ochre). Dots: assignment via available IHC. Open circles: due to lacking IHC, assignment by imputation via DST. Decision borders according to Eq. (11) are indicated by dotted lines at $\alpha_H = 0.5$, $\beta_H = 0.5$ and $\theta_H = 0.5$. Samples with same status assigned appear confined within the same respective tetrahedral space. Note that all points lie in one plane due to the definition ($\alpha_H + \beta_H + \theta_H = 1$, see SuppEq. 16) within a triangle (surrounded by heavy lines). (**B**) The triangle plane in panel (A) seen from a slightly different aspect, along a vector parallel to plane ($\alpha_H . \beta_H$), and perpendicular to the baseline of the triangle. Decision borders are more clearly visible in this aspect. In this projection the point ($\alpha_H = 1$, $\beta_H = 1$) in the foreground of panel (A) coincides with the baseline of the triangle, and so do the axes α_H and β_H. Note that they run in opposite directions. Hormone receptor positive samples lie exclusively in the lower right triangle, negative ones in the lower left whereas uncertain results in the kite-shaped area.

Discussion

Dempster Shafer decision theory is a very potent framework. Its strength lies in combining information from several sources about the same item (here exemplified by IHC, Gene and Co-Gene estimates for one receptor) and also evidences for several different items, interacting in a biological setting (here exemplified by estrogen and progesterone regarding precision therapy of breast cancer).

We have started from some best methodology available up to now (sODDS), i.e. receptor status assessment via IHC enhanced by gene expression. ODDS had already been shown to improve precision therapy[57], and here we demonstrate that further improvement is possible by application of DST. Since absolute truth of receptor status is unknown, the usefulness of our approach is demonstrated by disease free survival being severely degraded in patients flagged by DST as possibly wrongly diagnosed and treated. Explicit data on treatment are rare in the breast cancer studies downloaded from GEO. Hence we performed our calculations on the reasonable assumption that treatment was given according to the best knowledge available about receptor status.

The potency of DST is further underpinned by directly comparing both methods applied to the same dataset: DST flags about twice as many patients as 'uncertain' (153) as ODDS (82). Additional patients subjected to a re-evaluation of hormone receptor status will improve precision of therapy.

Formulae and results derived therefrom incorporate three assumptions which we think were reasonable:

To compute basic belief assignments for each patient, we obtained DST mass functions by logistic regression. It proved more stable than other possible methods, e.g. fitting bimodal distributions such as Gaussian mixtures, as done in a previous study[58].

For joining evidences from 2 genes for the same receptor we chose the Dempster ECR. For adding IHC evidence to those from gene expression, we opted for the Yager ECR.

To obtain upper limits $\hat{\alpha}, \hat{\beta}$ for belief we reasonably assumed that the fraction of positive results by virtue or chance equals the fraction of IHC-true positives (Eq. 3). Likewise we assumed that the ratio of uncertain positive and negative cases equals the ratio of positives and negatives (Eq. 2).

It is interesting to note that up to now clinical decisions follow the conventional (Cantor) non-exclusive OR-logics: 'If ER or PGR receptor estimates is/are positive, the patient is considered positive'. In part, the strength of DST originates from elaborate evidence joining: Simple and/or combinations (such as estrogen-or-progesterone receptor positivity) are fundamentally refined, since DST draws on *two* numbers (belief and plausibility) in each decision, instead of just a single number as conventional probability theory does. In addition, information from different sources is joined according to mathematical rules rather than 'clinical intuition'.

Joining evidences from both receptors in a more intricate way according to DST, allows for an intuitively convincing view on the classification, see Fig. 9. From the definition of $\theta_H = 1 - \alpha_H - \beta_H$ follows $\alpha_H + \beta_H + \theta_H = 1$ and, accordingly, DST beliefs ($\alpha_H, \beta_H, \theta_H$) of all samples lie in one single plane in 3-dimensions, cutting through the axes at $\alpha_H = 1$, $\beta_H = 1$ and $\theta_H = 1$, see panel A. Decision boundaries (dotted lines in Fig. 9) originate at 0.5 on each axis of evidence ($\alpha_H = 0.5$, $\beta_H = 0.5$) and confine tetrahedrons, each containing exclusively positive (red) or negative (blue) samples, respectively. These tetrahedrons appear as equilateral triangles in the view shown

in panel B. Uncertain samples (shown in ochre) appear within the pentahedron (panel A), which reduces to a kite-shaped area in the view shown in panel B.

While in this paper we have presented a very simple example to demonstrate feasibility and benefit of the method, possible applications are wide in scope. Pathology and laboratory medicine in many cases have to deliver crisp decisions even if measurements by complementary methods lack agreement. DST is the method of choice to handle such situations.

The final evidence obtained by DST also allows assigning weights (loss functions) to certain outcomes in order to model clinical consequences (Eq. 11). Suppose, for example, a false positive outcome entails much more detrimental consequences than a false negative does.

Several different methods are available for joining: The 'Dempster additivity rule', $\oplus_D$, works best with rather concordant variables. Needless to mention that correlation must not be too close to unity, as in this case the second variable would not add any significant information to the first one. As opposed, the 'Yager rule', $\oplus_Y$, lends itself to join evidences from rather contradicting sources.

OMICS data often comprise a considerable number of estimates, say counts for multiple fragments out of a gene, needing to be condensed to characterize the gene (activity) as a whole. A plethora of similar situations can rigorously be handled by DST. Even if some of these estimates should be (partly) contradictive, DST is an appropriate tool for consolidating. Formulae need to be adapted to the particular situation in question.

We think that this work may help pave the way for decision theory entering OMICS data science.

References

1. Toss, A. & Cristofanilli, M. Molecular characterization and targeted therapeutic approaches in breast cancer. *Breast Cancer Res.* **17**, 60 (2015).
2. Dowsett, M. *et al.* Comparison of PAM50 risk of recurrence score with oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. *J. Clin. Oncol.* **31**, 2783–2790 (2013).
3. Prat, A., Ellis, M. J. & Perou, C. M. Practical implications of gene-expression-based assays for breast oncologists. *Nat. Rev. Clin. Oncol.* **9**, 48–57 (2011).
4. Desmedt, C. *et al.* Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clin. Cancer. Res.* **14**, 5158–5165 (2008).
5. 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* **11**, 143 (2011).
6. Jezequel, P. *et al.* Gene-expression molecular subtyping of triple-negative breast cancer tumours: importance of immune response. *Breast Cancer Res.* **17**, 43 (2015).
7. Burstein, M. D. *et al.* Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. *Clin. Cancer. Res.* **21**, 1688–1698 (2015).
8. Wu, G. & Stein, L. A network module-based method for identifying cancer prognostic signatures. *Genome Biol.* **13**, R112 (2012).
9. Liu, R., Guo, C. X. & Zhou, H. H. Network-based approach to identify prognostic biomarkers for estrogen receptorGÇópositive breast cancer treatment with tamoxifen. *Cancer Biol. Ther.* **16**, 317–324 (2015).
10. Korde, L. A. *et al.* Gene expression pathway analysis to predict response to neoadjuvant docetaxel and capecitabine for breast cancer. *Breast Cancer Res. Treat.* **119**, 685–699 (2010).
11. Clarke, C. *et al.* Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis. *Carcinogenesis* **34**, 2300–2308 (2013).
12. Aswad, L. *et al.* Genome and transcriptome delineation of two major oncogenic pathways governing invasive ductal breast cancer development. *Oncotarget* **6**, 36652–36674 (2015).
13. Filipits, M. *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.* **17**, 6012–6020 (2011).
14. Zhao, X. *et al.* Systematic assessment of prognostic gene signatures for breast cancer shows distinct influence of time and ER status. *BMC. Cancer* **14**, 211 (2014).
15. van't Veer, L. J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536 (2002).
16. van de Vijver, M. J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347**, 1999–2009 (2002).
17. Paik, S. *et al.* A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.* **351**, 2817–2826 (2004).
18. Lu, X. *et al.* Predicting features of breast cancer with gene expression patterns. *Breast Cancer Res. Treat.* **108**, 191–201 (2008).
19. Budczies, J. *et al.* Genome-wide gene expression profiling of formalin-fixed paraffin-embedded breast cancer core biopsies using microarrays. *J Histochem. Cytochem* **59**, 146–157 (2011).
20. Lin, C. Y. *et al.* Discovery of estrogen receptor a target genes and response elements in breast tumor cells. *Genome Biol.* **5**, R66 (2004).
21. Liu, M. C. *et al.* PAM50 gene signatures and breast cancer prognosis with adjuvant anthracycline- and taxane-based chemotherapy: correlative analysis of C9741 (Alliance). *Npj Breast Cancer* **2**, 15023 (2016).
22. Prat, A. *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.* **20**, 511–521 (2014).
23. Wishart, G. C. *et al.* PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. *Breast Cancer Res.* **12**, R1 (2010).
24. Desmedt, C. *et al.* The gene expression grade index: a potential predictor of relapse for endocrine-treated breast cancer patients in the BIG 1GÇô98 trial. *BMC Med. Genom.* **2**, 40–40 (2009).
25. Metzger-Filho, O. *et al.* Genomic grade index (GGI): feasibility in routine practice and impact on treatment decisions in early breast cancer. *PLoS ONE* **8**, e66848 (2013).
26. Huang, C. C. *et al.* Concurrent gene signatures for han chinese breast cancers. *PLoS ONE* **8**, e76421 (2013).
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.* **53**, 292–301 (2000).

28. Wolff, A. C. 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.* **25**, 118–145 (2006).

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

30. Harris, L. N. 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.* **34**, 1134–1150 (2016).

31. Singer, C. F. et al. Pathological complete response to neoadjuvant trastuzumab is dependent on HER2/CEP17 Ratio in HER2-amplified early breast cancer. *Clin. Cancer. Res.* **23**, 3676–3683 (2017).

32. Harbeck, N. & Gnant, M. Breast cancer. *Lancet* **389**, 1134–1150 (2016).

33. Wirapati, P. 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.* **10**, R65 (2008).

34. Hudis, C. A. et al. Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: the STEEP system. *J. Clin. Oncol.* **25**, 2127–2132 (2007).

35. Regan, M. M. et al. Re-evaluating adjuvant breast cancer trials: assessing hormone receptor status by immunohistochemical versus extraction assays. *J. Natl. Cancer Inst.* **98**, 1571–1581 (2006).

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* **117**, 1575–1582 (2011).

37. Bartlett, J. M. et al. A UK NEQAS ISH multicenter ring study using the Ventana HER2 dual-color ISH assay. *Am. J. Clin. Pathol.* **135**, 157–162 (2011).

38. Lee, M., Lee, C. S. & Tan, P. H. Hormone receptor expression in breast cancer: postanalytical issues. *J. Clin. Pathol.* **66**, 478–484 (2013).

39. 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.* **6**, 195–197 (2010).

40. Wells, C. A. et al. Consistency of staining and reporting of oestrogen receptor immunocytochemistry within the European Union: an inter-laboratory study. *Virchows Arch.* **445**, 119–128 (2004).

41. Laas, E. et al. Low concordance between gene expression signatures in ER positive HER2 negative breast carcinoma could impair their clinical application. *PLoS ONE* **11**, e0148957 (2016).

42. Rakha, E. A. et al. Updated UK Recommendations for HER2 assessment in breast cancer. *J. Clin. Pathol.* **68**, 93–99 (2015).

43. Allred, D. C. et al. NCCN task force report: estrogen receptor and progesterone receptor testing in breast cancer by immunohistochemistry. *J. Natl. Compr. Canc. Netw* **7**(Suppl 6), S1–S21 (2009).

44. Li, Q. et al. Minimising immunohistochemical false negative ER classification using a complementary 23 gene expression signature of ER status. *PLoS ONE* **5**, e15031 (2010).

45. 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.* **14**, 5959–5966 (2008).

46. Wu, Z. A review of statistical methods for preprocessing oligonucleotide microarrays. *Stat. Methods Med. Res.* **18**, 533–541 (2009).

47. Wu, Z. & Irizarry, R. A. Preprocessing of oligonucleotide array data. *Nat. Biotechnol.* **22**, 656–658 (2004).

48. 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.* **99**, 909–917 (2004).

49. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, Article 17 (2005).

50. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

51. Wu, Z. & Irizarry, R. A. A statistical framework for the analysis of microarray probe-level data. *Ann. Appl. Stat.* **1**, 333–357 (2007).

52. 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* **19**, 185–193 (2003).

53. Chen, X. et al. TNBCtype: a subtyping tool for triple-negative breast cancer. *Cancer Inform.* **11**, 147–156 (2012).

54. Gong, Y. et al. Determination of oestrogen-receptor status and ERBB2 status of breast carcinoma: a gene-expression profiling study. *Lancet Oncol.* **8**, 203–211 (2007).

55. Bergqvist, J. et al. Quantitative real-time PCR analysis and microarray-based RNA expression of HER2 in relation to outcome. *Ann. Oncol.* **18**, 845–850 (2007).

56. 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.* **50**, 981–990 (2012).

57. Kenn, M. et al. Co-expressed genes enhance precision of receptor status identification in breast cancer patients. *Breast Cancer Res. Treat.* **172**, 313–326. https://doi.org/10.1007/s10549-018-4920-x (2018).

58. Kenn, M. et al. Gene expression information improves reliability of receptor status in breast cancer patients. *Oncotarget* **8**, 77341–77359 (2017).

59. Gordon, J. & Shortliffe, E. H. in *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project* (eds. Buchanan, B. G. & Shortliffe, E. H) Chap. 13, 832–838 (Addison-Wesley Publishing Company, 1984).

60. Högger, A. Dempster Shafer Sensor Fusion for Autonomously Driving Vehicles : Association Free Tracking of Dynamic Objects, diplomar thesis, KTH Royal Institut of Technology School of Electrical Engineering, Sweden, http://kth.diva-portal.org/smash/get/diva2:931627/FULLTEXT01.pdf (2016).

61. 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 *2009 International Conference on Networking, Sensing and Control*, pp. 897–902 (2009).

62. Jugade, S. C. & Victorino, A. C. Grid based estimation of decision uncertainty of autonomous driving systems using belief function theory. *IFAC-PapersOnLine* **51**, 261–266. https://doi.org/10.1016/j.ifacol.2018.07.043 (2018).

63. 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.* **71**, 392–401. https://doi.org/10.1016/j.ast.2017.09.040 (2017).

64. 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. Failure Anal.* **18**, 2084–2092. https://doi.org/10.1016/j.engfailanal.2011.06.014 (2011).

65. 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.* **8**, 593–607. https://doi.org/10.1109/TIFS.2013.2248727 (2013).

66. Chandana, S., Leung, H. & Trpkov, K. Staging of prostate cancer using automatic feature selection, sampling and Dempster-Shafer fusion. *Cancer Inform.* **7**, 57–73 (2009).

67. 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* **1**, 170–175 (2006).

68. van Vliet, M. H. et al. Pooling breast cancer datasets has a synergetic effect on classification performance and improves signature stability. *BMC Genom.* **9**, 375 (2008).

69. Rung, J. & Brazma, A. Reuse of public genome-wide gene expression data. *Nat. Rev. Genet.* **14**, 89–99 (2013).

70. Edgar, R. & Barrett, T. NCBI GEO standards and services for microarray data. *Nat. Biotechnol.* **24**, 1471–1472 (2006).

71. Barrett, T. & Edgar, R. Mining microarray data at NCBI's gene expression omnibus (GEO). *Methods Mol. Biol* **338**, 175–190 (2006).
72. Irizarry, R. A. *et al.* Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31**, e15 (2003).
73. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307–315 (2004).
74. Stafford, P. *Methods in Microarray Normalization* (CRC Press, Boca Raton, 2008).
75. McCall, M. N., Jaffee, H. A. & Irizarry, R. A. fRMA ST: frozen robust multiarray analysis for Affymetrix Exon and Gene ST arrays. *Bioinformatics (Oxford, England)* **28**, 3153–3154. https://doi.org/10.1093/bioinformatics/bts588 (2012).
76. Bolstad, B. *Background and Normalization: Investigating the effects of preprocessing on gene expression estimates*, http://bmbolstad.com/stuff/BAUGM.pdf (2017).
77. Kenn, M. *et al.* Microarray normalization revisited for reproducible breast cancer biomarkers. *Biomed. Res. Int.* **2020**, 1363827. https://doi.org/10.1155/2020/1363827 (2020).
78. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264 (2003).
79. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* 2nd edn. (Springer, Berlin, 2009).
80. Gendoo, D. M. *et al.* Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer. *Bioinformatics* **32**, 1097–1099 (2016).
81. McCullagh, P. & Nelder, J. A. *in Monographs on Statistics and Applied Probability Volume 37 Generalized Linear Models* (Chapman and Hall/CRC, London, 1989).
82. Shoyaib, M., Abdullah-Al-Wadud, M. & Chae, O. A skin detection approach based on the Dempster-Shafer theory of evidence. *Int. J. Approx. Reason.* **53**, 636–659 (2012).
83. Yang, J. B. & Xu, D. L. Evidential reasoning rule for evidence combination. *Artif. Intell.* **205**, 1–29 (2013).

Acknowledgements

Author contributions

Conceptualization: W.S., H.K., C.S.; Methodology and software: M.K.; Formal analysis: M.K.; Data curation: M.K. and M.C.; Writing—original draft preparation: W.S.; Writing—review and editing: D.C.C., R.K., M.C. and W.S.; Supervision: W.S.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at (https://doi.org/10.1038/s41598-021-82418-7).

Correspondence and requests for materials should be addressed to W.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.