Discovery of pharmaceutically-targetable pathways and prediction of survivorship for pneumonia and sepsis patients from the view point of ensemble gene noise

Tristan de Jong[1,4], Victor Guryev[1,4], Yury M. Moshkin[2,3,4]

[1] European Research Institute for the Biology of Ageing, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands,

[2] Federal Research Centre, Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia,

[3] Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia,

[4] Gene Learning Association, Geneva, Switzerland

Correspondence: yury.moshkin@gmail.com; v.guryev@umcg.nl

Abstract

Finding novel biomarkers for human pathologies and predicting clinical outcomes for patients is rather challenging. This stems from the heterogenous response of individuals to disease which is also reflected in the inter-individual variability of gene expression responses. This in turn obscures differential gene expression analysis (DGE). In the midst of the COVID-19 pandemic, we wondered whether an alternative to DGE approaches could be applied to dissect the molecular nature of a host-response to infection exemplified here by an analysis of H1N1 influenza, community/hospital acquired pneumonia (CAP) and sepsis. To this end, we turned to the analysis of ensemble gene noise. Ensemble gene noise, as we defined it here, represents a variance within an individual for a collection of genes encoding for either members of known biological pathways or subunits of annotated protein complexes. From the law of total variance, ensemble gene noise depends on the stoichiometry of the ensemble genes' expression and on their average noise (variance). Thus, rather than focusing on specific genes, ensemble gene noise allows for the holistic identification and interpretation of gene expression disbalance on the level of gene networks and systems. Comparing H1N1, CAP and sepsis patients we spotted common disturbances in a number of pathways/protein complexes relevant to the sepsis pathology which lead to an increase in the ensemble gene noise. Among others, these include mitochondrial respiratory chain complex I and peroxisomes which could be readily targeted for adjuvant treatment by methylene blue and 4-phenylbutyrate respectively. Finally, we showed that ensemble gene noise could be successfully applied for the prediction of clinical outcome, namely mortality, of CAP and sepsis patients. Thus, we conclude that ensemble gene noise represents a promising approach for the investigation of molecular mechanisms of a pathology through a prism of alterations in coherent expression of gene circuits.

Introduction

Both viral and bacterial pneumonia may lead to a life-threatening condition, namely sepsis. Most notable cases, in the public perception, include pandemic viral infections, such as the 2009 swine flu pandemic caused by H1N1 [1] and more recently, the 2019 coronavirus disease (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [2]. Like with any other annual severe acute respiratory infections (SARI), these pandemics resulted in a significant raise in patients with sepsis at intensive care units[3, 4]. Sepsis is a complex reaction of the host (human) to a systemic infection (viral or bacterial) often resulting in septic shock or death [5-7]. A problem of sepsis treatment, the prediction of patients' clinical outcomes and the risks of mortality relates to the highly heterogenous nature of sepsis [8]. Thus, despite recent progress in identification of molecular biomarkers for sepsis [8-15], treatment remains mainly non-curative and clinical outcomes are mostly inferred from clinical signs [5].

A canonical approach for the identification of disease biomarkers and their potential therapeutic targets relies on differential gene expression (DGE) analysis either on RNA or protein levels. This stems from a classical gene regulation Jacob-Monod model, which implies a specific gene expression response (up- or down-regulation) to a specific signal (see recent perspective on historical origins of the model in [16]. However, gene expression is a stochastic process and cellular responses to signals often trigger a cascade of changes in gene expression, making it difficult to discover specific targets and biomarkers for a disease.

The stochastic nature of gene expression implies a natural variation in RNA and protein copy numbers [17]. According to the fluctuation-response relationship [18, 19], an amount of gene expression response to a signal (fluctuation) is proportional to its variance (or squared biological coefficient of variation – bcv^2) for log-scaled values of RNA copy number [20]. Consequently, statistical inference of differentially expressed genes will be biased towards genes with high variance (bcv^2) (Figure S1). This leads to a set of intrinsic problems with DGE analysis. 1) genes with increased variability in expression will strongly respond to any cellular signal aimed at them. However, these genes may not necessarily be causative for a diseased state. Even under normal circumstances they exhibit large fluctuations and, thus, are loose-regulated. 2) In contrast, genes with a low variability will respond only modestly, but these genes are tight-regulated and any fluctuations in their expression might be causative for a diseased state.

Upon calling significantly changed genes, to make biological sense, these genes are mapped to known biological pathways, such as GO or KEGG [21, 22], or to subunits of protein complexes annotated by CORUM or other interaction databases [23]. Thus, a second statistical test is required, namely gene set enrichment analysis (GSEA). However, this is not without its own caveats. The major one is that GSEA depends on the statistical inference of DGE and DGE cut-offs [24, 25]. As a result, biological interpretations from DGE might be drastically affected by pitfalls arising from the fluctuation-response relationship, DGE thresholding and the choice of statistical approach for GSEA.

To circumvent this, we reasoned that 1) genes do not function in isolation, but rather act as ensembles representing biological pathways and/or subunits of protein complexes. 2) The normal function of a biological pathway or protein complex requires a regulated (balanced) expression of the whole gene ensemble. 3) Any alterations in the expression of a gene ensemble might be causative for a disease or predictive for clinical outcome. To infer the alterations in gene ensembles expression we turned to the estimation of their variances (ensemble gene noise) from whole blood gene expression profiles of individuals under normal and pathological conditions. From the total law of variance, ensemble gene noise ($\mathrm{Var}[G]$) sums from the variance of ensemble genes' means and ($\mathrm{Var}\big[\mathrm{E}[G|g]\big]$), and the expectation of ensemble genes variances ($\mathrm{E}\big[\mathrm{Var}[G|g]\big]$) (Figure S2). Thus, the ensemble gene noise estimates both: 1) changes in stoichiometries of genes encoding either a biological pathway or protein complex subunits and 2) changes in mean gene expression variability for genes in ensemble.

From the whole blood expression profiles of patients under intensive care treatment we estimated how ensemble gene noise corresponds to a pathological state, such as sepsis, community/hospital acquired pneumonia (CAP) or viral H1N1 pneumonia (H1N1). From this analysis we identified a number of pathways for which ensemble gene noise associated positively with an individual health/disease state treated as an ordinal variable (healthy < early H1N1 phase < late H1N1 phase and healthy < sepsis/CAP survived < sepsis/CAP deceased patients). Finally, we identified pathways and complexes where deregulation is associated with a poor prognosis and predicted the clinical outcome (survival/mortality) for CAP/sepsis patients based on ensemble gene noise with high accuracy. We concluded that the ensemble gene noise provides a powerful tool for the discovery of systemic disease biomarkers, pharmaceutically targetable pathways and the prediction of a disease clinical outcome.

4

Results

Mean and variance gene expression response to infection and sepsis

Sepsis is thought to trigger a plethora of heterogenous host responses to a systemic infection [5, 8]. We reasoned that this heterogeneity might be reflected in the inter-individual gene expression variability (standard deviation - σ or variance - σ^2). Considering that a) RNA copy number is a mixed Poisson (*e.g.* negative binomial) random variable [26] and that b) log-transformed microarray hybridization signal intensities correlate with log-transformed RNA-seq copy numbers [27]. It is easy to show that the variance of log gene expression approximates the biological coefficient of variation (bcv^2) [20]. From the first-order Taylor expansion for variance: $\sigma_Y^2 \approx \frac{\sigma_X^2}{\bar{X}^2} = cv_X^2$, where $Y = log(X)$ is the log gene expression. The mixed Poisson random variable, $cv_X^2 = \frac{1}{\mu_X} + bcv^2$, where bcv^2, also known as the overdispersion parameter, is independent of mean gene expression (μ_X). Thus, for $\mu_X \gg 1$ (for genes with a large mean RNA copy number), $\sigma_Y^2 \approx bcv^2$. In other words, by estimating the inter-individual log gene expression variabilities from either microarray signal intensities or RNA-seq counts we can infer approximately the biological coefficients of variations for genes' RNA copy numbers.

We estimated the dispersions for whole blood log gene expressions in CAP and sepsis patients (8826 genes), and H1N1 infected patients (7240 genes) from the two data sets GSE65682 and GSE21802 respectively (for a detailed description of cohorts see original studies and Methods) [8, 11, 28]. For CAP and sepsis patients we also accounted for age, including it as a random variable in the Generalized Additive Model for Location, Scale and Shape (GAMLSS) [29], see Methods. On average, the dispersions in log gene expressions in CAP, sepsis and H1N1 patients were significantly higher as compared to healthy individuals (Figure 1A). To that, for CAP patents' dispersions in log gene expressions were significantly higher for deceased patients as compared to those survived. Likewise, for H1N1 patients, dispersions in log gene expressions further increased in the late phase of infection (Figure 1A). For sepsis patients, on average dispersions in log gene expressions were comparable between survived and deceased patients for all analysed genes (Figure 1A). However, for genes for which dispersions changed significantly between healthy individuals and sepsis patients (Bonferroni adjusted p ≤ 0.05), their dispersions on average were higher in the deceased patients as compared to the survived (p < 0.001). Together, these suggest that host response to infection increases the biological

coefficients of variations in genes' RNA copy numbers (as $\sigma^2 \approx bcv^2$) and substantiates heterogeneity in the pathogenesis of sepsis [8] from the gene expression perspective.

Because of the fluctuation-response relationship [18], absolute changes in the mean log gene expressions ($|\Delta\mu|$) in response to infection (CAP, H1N1) and sepsis correlated significantly with the variances of the log gene expressions (Figure 1B). Interestingly, we also noted significant correlations between the absolute changes in inter-individual gene expression variabilities ($|\Delta\sigma^2|$) and the variances of log genes expressions (Figure 1C). Consequently, $|\Delta\mu|$ and $|\Delta\sigma^2|$ were also correlated (Figure 1D). Thus, we conclude that H1N1, CAP and sepsis result in coordinated changes in both the mean and heterogeneity of the expression of genes and that magnitudes of these changes depend on genes' biological coefficients of variation.

Ensemble gene noise response to infection and sepsis

Both the mean and variance relate to population (inter-individual) statistics reflecting distinct aspects of gene regulation. Changes in means fit the classical DGE view on gene response to a pathology and other biological processes, while changes in variances yield a view on heterogeneity of gene response. However, as we noted before (Figure 1 and S1), statistical inference of these changes is biased towards higher a significance for genes with a high biological coefficient of variation. Although changes in RNA copy number can serve in practical applications for diagnostics of a disease and clinical outcomes, inter-individual variability cannot be used for diagnosis. At the same time, stochastic fluctuations in gene expression remain attractive for the dissection of novel molecular mechanisms of a pathology. Therefore, we expect that estimation of ensemble gene noise may provide additional benefits for diagnostics by quantifying fluctuations, while being informative for personalized treatment.

We define ensemble gene noise as the variance of log-transformed, normalized expression levels for a collection of genes $G = (g_1, \dots, g_i)$ encoding for either proteins of a pathway or subunits of a protein complex. To this end, we mapped genes to the KEGG-annotated pathways and the CORUM-annotated protein complexes [22, 23]. From the law of total variance: $\text{Var}[G] = \text{E}[\text{Var}[G|g]] + \text{Var}[\text{E}[G|g]]$, ensemble gene noise depends on the variability in expression of genes in ensemble ($\text{E}[\text{Var}[G|g]]$) and on their stoichiometry ($\text{Var}[\text{E}[G|g]]$) (Figure S1). Thus, the simple estimation of the variances ($\text{Var}[G]$) of gene ensembles for each

individual might reflect alterations in function of biological pathways and protein complexes on the level of stoichiometry and gene noise.

We, then, correlated $Var[G]$ for ensembles with H1N1, CAP and sepsis disease states. For H1N1 viral infection, disease state can be clearly ranked: non-infected (healthy) < early phase < late phase of infection, thus it represents an ordinal variable [28]. For CAP and sepsis patients, we assumed that a condition of the deceased patients was worse than that of the survived. We considered that healthy < survived < deceased can also be represented as ordinal disease state variable. Circumstantially, this is supported by distinct blood gene expression endotypes [8] and an increased gene expression heterogeneity (Figure 1A). Kendall rank correlation identified a number of pathways and protein complexes for which ensemble gene noise was positively and significantly associated with the disease state in H1N1 (FDR ≤ 0.05), and CAP and sepsis patients (Bonferroni-adjusted p ≤ 0.05) (Figure 2A). None of the pathways or gene complexes were negatively associated with the disease state at the specified significance thresholds. We used different p value adjustment procedures (FDR – less conservative, and Bonferroni – more conservative) for H1N1, CAP and sepsis patents due to the large differences in sample sizes (number of patients) between these data sets.

Out of all gene ensembles, 13 of them proved to be consistent and correlated to the increased disease state in ensemble gene noise in all three disease conditions (Figure 2A, B, Table S1A). Most of these gene ensembles (pathways) are known to be involved in the pathology of sepsis through multiple experimental evidences (Table 1), thus substantiating a power of ensemble gene noise analysis. However, ensemble gene noise yields novel insights into the molecular mechanisms of sepsis (H1N1, CAP or other-causes of sepsis) by suggesting a holistic mis-regulation in stoichiometry and gene noise for these gene ensembles.

We also identified 5 gene ensembles for which ensemble gene noise was positively and significantly correlated with the disease state in H1N1 and CAP patients (Figure 2A, Table S1B). However, ensemble gene noise for these pathways was also significantly increased in sepsis patients (t-test, Bonferroni adjusted p < 0.01) despite insignificant rank correlation. To that, some of these pathways can be implicated in the pathology of sepsis (Table 1). Two of these ensembles were represented by genes encoding mitochondrial respiratory chain complex I (Complex I) (Figure 2C). From the point of view of ensemble gene noise this suggests an altered stoichiometry and gene noise in the expression of the subunits of the Complex I which, as a result, might lead to its improper assembly and function in H1N1, CAP and sepsis patients.

Indeed, it has been established that the activity of the Complex I is decreased and correlates with the severity of sepsis [30]. Complex I is the first set of enzymes of the respiratory chain and it is the entry point for most electrons into the electron transport chain [31]. Interestingly, however, in case of the Complex I inhibition or deregulation, methylene blue (MB) can bypass it by acting as alternative redox mediator in the electron transport chain, thus, restoring mitochondrial respiration [32, 33] (Figure 2D). MB is also considered to be a promising therapeutic in treatment of septic shock [34, 35]. Thus, ensemble gene noise might provide a simple yet powerful explanatory shortcut, from the expression of thousands of genes to the function of gene ensembles and possible pharmaceutical targets.

Predicting clinical outcome for CAP and sepsis patients from the ensemble gene noise

Treatment of sepsis is challenging and mortality rates among sepsis patients are high. Yet, prediction of clinical outcome is also challenging due to heterogeneity in the pathology [8] and gene expression (Figure 1A). Recently, Molecular Diagnosis and Risk Stratification of Sepsis (MARS) consortium identified the Mars1 gene expression endotype which was significantly associated with acute (28-day) mortality, however, for other endotypes Mars2-4 poorly discriminated between the survival and mortality of patients [8]. Thus, we wondered whether the clinical outcome (mortality) could be predicted from the ensemble gene noise.

To this end, we trained binary logistic gradient boosted regression tree models using survival and acute mortality as a binary response variable for clinical outcome and patients' age and blood ensemble gene noise as models' features. The models were trained with XGBoost [36]. The CAP and sepsis patients were split into discovery (263 patients: 105 CAP and 158 sepsis patients) and validation (216 patients: 78 CAP and 138 sepsis patients) cohorts following GSE65682 annotation [8]. Within the cohorts the mortality rates were 26.2% for CAP and sepsis patients (23.8% for CAP and 27.8% for sepsis patients) in the discovery cohort, and 20.8% for CAP and sepsis patients (19.2% for CAP and 21.7% for sepsis patients) in the validation cohort.

Overall, class-imbalance, noise due to the inter-individual heterogeneity and high-dimensionality of model features are among the major problems of machine learning [37]. In part, ensemble gene noise leads to a reduction in inter-individual variability (Figure S3) and in dimensionality as model features are represented not by individual genes, but by collections of genes. Nonetheless, we further reduced the number of ensemble gene noise features in models

8

by t-test feature selection. For this, we compared gene ensembles noise between survived and deceased patients in the discovery cohorts. The p-value cut-offs for the model features were selected based on maximization of models' training accuracy (see Methods). XGBoost hyper-tuning parameters: learning rate, complexity, depth, *etc.* were optimized based on the cross-validation. To avoid overfitting, we used early epoch stopping, which was estimated from the test fold of the discovery cohort (see Methods). Because of the class-imbalance, AUC (area under the receiver operating characteristic (ROC) curves) was used to evaluate the model performance. The validation cohorts were hidden from the feature selection and training.

Figures 3A and 3B show model scores and ROC curves for the model, predicting mortality/survival for the CAP and sepsis patients in the discovery and validation cohorts. AUCs for the discovery and validation cohorts were 0.871 and 0.707 respectively, suggesting a reasonable accuracy of the model. However, from the model scores, and evaluation of the model specificity/sensitivity it appears that the model is biased towards the prediction of major class (survived) (Figure 3A, Table 2 and Table S2A). Thus, class prediction balanced accuracies (bACC = Specificity/2 + Sensitivity/2) were 0.799 and 0.701 for the discovery and validation cohorts respectively. Nonetheless, the survival probability for patients predicted to have a high risk of mortality was significantly lower than the survival probability of patients predicted to have low risk of mortality in both discovery and validation cohorts. To that, our model better predicts the risks of mortality as compared to the Mars1 endotype inferred from the log gene expression unsupervised learning (Figure 3C) [8]. Potentially, this could be due to a lower inter-individual variability of gene ensembles noise as compared to log gene expression (Figure S3).

In an attempt to increase the prediction accuracy, we trained to separated gradient boosted tree models for CAP (Figure 4) and sepsis (Figure 5) patients. Indeed, in both cases the accuracy of the prediction of the minor class (deceased patients) increased (Table 2 and S2) in both discovery and validation cohorts. Likewise, AUCs for the validation cohorts were also higher as compared to the model predicting mortality for both (CAP and sepsis) type of patients (compare Figure 4B and 5B with Figure 3B). To that, differences in AUCs between discovery and validation cohorts were lower for the models predicting mortality separately for CAP and sepsis patients as for the model trained on both type of patients. This was especially evident for the model predicting mortality for the CAP patients (Figure 4B). Thus, knowing the cause of sepsis improves the prediction accuracy of the models.

9

Finally, it has to be noted that both the feature selection and gradient boosted regression trees allow for the ranking of the model features' importance (Figures 3D, 4D, 5D and Table S3). First, it turned out that a patients' age does not noticeably contribute to the prediction of mortality in CAP patients and it ranks low in the prediction of mortality of sepsis patients. Second, high ranking gene ensembles (pathways) could be immediately associated with host response to infection and, thus, pathology of the sepsis. These include legionellosis (a pathway responsible for atypical pneumonia caused by *Legionella* bacteria), epithelial cell signalling in *Helicobacter pylori* infection and leishmaniasis, and imbalances in these pathways either caused by corresponding infections or immune activation could lead to the sepsis [38-40]. To that, ensemble gene noise in immune pathways, such as rheumatoid arthritis and primary immunodeficiency, contribute to the prediction of clinical outcome in sepsis patients (Figure 3D, 5D and Table S3). Thus, we conclude that the ensemble gene noise uncovers novel approaches and insights to the discovery of biomarkers, prediction of clinical outcome and to the molecular mechanisms of a pathology from the point of view of imbalances in stoichiometry and gene noise of expression in gene ensembles.

Discussion

Here we attempted a dissection of molecular mechanisms of human pathology, exemplified by H1N1 infection, CAP and sepsis, through a prism of ensemble gene noise. Unlike classical DGE, ensemble gene noise allows for the identification of imbalances in the expression of entire gene circuits, rather than individual genes on the level of stoichiometry and gene noise. This approach offers an alternative, but non-mutually exclusive to the DGE interpretation of a molecular basis of disease and both have their own strengths and weaknesses.

We noted in the introduction that due to a fluctuation-response a statistical inference of DGE might be biased towards genes with a high inter-individual variability, *i.e.* "noisy" genes (Figure S1) [18, 19]. However, the same applies to ensemble gene noise (Figure S4). This imposes a certain problem to the interpretation of both DGE and ensemble gene noise. On one hand, it can be suggested that large deviations in expression of genes and ensembles, which are naturally prone to high fluctuations, might not be causative for a disease, as an organism is already adapted to such variations. On the other hand, these genes/ensembles themselves might play an important adaptive role [41] and their over-response could lead to a disease. At the moment it seems difficult to come to a resolution between these two possibilities, but they

should be considered, specifically in identification of pharmaceutical targets: genes or gene ensembles (pathways, protein complexes).

As compared to DGE, ensemble gene noise provides a holistic interpretation to mis-regulation in gene expression under pathologic or other conditions. As it operates on the level of gene ensembles it does not require gene set enrichment analysis (GSEA), thus it circumvents potential pitfalls of GSEA associated with the cut-off problem of DGE [24, 25]. As any gene expression analysis ensemble gene noise relies on the quality and completeness of pathways and the protein complexes' annotation. Finally, we noted that inter-individual variability of ensemble gene noise is significantly less than that of individual gene expression (Figure S3). This, in turn, might improve the accuracy of diagnostic and clinical outcome models. Though it might come at the expense of less features being available for the selection and training of models. At the same time, in future studies, both DGE and ensemble gene noise could be combined.

In this study we applied the concept of ensemble gene noise to the analysis of critically ill H1N1, CAP and sepsis patients [8, 28]. We noted a large-scale gene response in two dimensions: on the level of mean gene expression and on the level of variance (inter-individual variability). Interestingly, both responses were correlated (Figure 1D) and both were dependent on gene variance suggesting that the fluctuation-response might drive changes in these two parameters of gene expression co-ordinately [18]. In all three cases (H1N1, CAP and sepsis), inter-individual variability was increased for a bulk of the genes. Consequently, we only identified pathways or gene complexes for which ensemble gene noise was significantly increased for H1N1, CAP and sepsis patients as compared to healthy individuals. This suggests that inter-individual gene expression variability is a prominent driver of ensemble gene noise in these patients.

Because viral/bacterial infections and sepsis result in overwhelming gene expression response, it is difficult to identify a reasonably small set of either genes or gene ensembles for biological interpretation. Thus, we only focused on the pathways (protein complexes) for which ensemble gene noise increased in all three cases and correlated these with a disease state (Figure 2A). From this intersection we inferred 13 pathways most of which have been previously implicated in sepsis (Table 1). To that, 5 pathways (protein complexes) showed significant association of ensemble gene noise with H1N1 infection phase and CAP disease state and for which ensemble gene noise also increased significantly in sepsis patients (Figure 2A). Potentially, these

pathways could be targeted for adjuvant treatment of sepsis. Especially, we consider mitochondrial respiratory chain complex I (Complex I) (Figure 2D) and peroxisome promising for pharmaceutical targeting. Increased ensemble gene noise for the Complex I would imply either altered stoichiometry, or increased gene expression noise for genes encoding subunits of the Complex I or both. As a result, this might lead to improper assembly of the Complex I and affecting its function. The impaired Complex I function can be bypassed by an alternative redox mediator, such as methylene blue [32, 33]. To that, methylene blue is a selective inhibitor of the nitric oxide–cyclic guanosine monophosphate (NO–cGMP) pathway [35] and increased NO levels is a hallmark of sepsis [42]. Some clinical studies have already indicated a beneficial role of methylene blue in the treatment of sepsis [34, 35]. Similar to mitochondrial respiration, peroxisomes also play an important role in the pathology of sepsis as the dysfunction of peroxisomes results in oxidative stress [43]. Again, an increased ensemble gene noise for peroxisome pathway indicates a potential mechanism for such dysfunction in H1N1, CAP and sepsis patients. Potentially peroxisome biogenesis could be restored by 4-phenylbutyrate and there several studies indicating its positive role in treatment of sepsis [44, 45]. Considering future directions, it could be proposed that search for epigenetic modulators of ensemble gene noise might represent a novel pharmaceutical avenue for adjuvant treatments of sepsis.

Finally, we explored the possibility to use ensemble gene noise in the prediction of clinical outcomes. Previously some promising biomarkers and gene expression endotypes associated with septic shock and mortality have been identified based on DGE analysis [8, 9]. However, as already mentioned, ensemble gene noise looks at gene expression from a different, yet complementary, angle, thus enabling the identification of novel pathways and biomarkers for sepsis and other diseases. To that, models predicting pathology based on ensemble gene noise could potentially be more robust, as inter-individual variability for ensemble gene noise is lower than that for log gene expression (Figure S3). Furthermore, Gradient boosted regression tree models trained on CAP and sepsis patients to predict their mortality had a good accuracy on validation cohort (Figure 3, Table 2). These outperformed predictions based on the Mars1 gene expression endotype, which was shown to associate with a poor prognosis [8], both on the discovery and validation cohorts (Figure 3C). Interestingly, some ensemble gene noise features selected statistically for the models predicting mortality in both CAP/sepsis-, CAP- and sepsis- patients couldimmediately be related to the host's response to infection. For example, increases in ensemble gene noise in legionellosis, epithelial cell signalling in

12

Helicobacter pylori infection and leishmaniasis pathways could potentially serve as biomarkers of sepsis and its outcome.

In conclusion, here we showed a potential of ensemble gene noise in the biological interpretation of a disease, the identification of pharmaceutically targetable pathways, novel biomarkers, and the prediction of clinical outcome. Together, we believe that ensemble gene noise analysis could be broadly applied alongside with DGE to dissect molecular mechanisms of the pathology in two complementary dimensions: in Jacob-Monod dimension of specific gene regulation and in a novel dimension of holistic gene circuit regulation.

Methods

Data resources and processing

GSE65682 Affymetrix Human Genome U219 Array whole blood gene expression profiles were used for the analysis of community/hospital acquired pneumonia (CAP) and sepsis patients [8, 11]. In brief, the cohort consisted of 42 healthy individuals (24 males, 18 females), 183 CAP patients (111 males, 72 females) and 296 sepsis patients (161 males, 135 females). The mean age of CAP (61.5 ± 1.2) and sepsis (60.6 ± 0.8) patients did not differ significantly ($t(350.37) = 0.59$, p = 0.56), however healthy individuals were significantly younger as compared to CAP ($t(54.0) = 4.7$, p < 0.001) and sepsis ($t(47.2) = 4.6$, p < 0.001) patients. Out of 183 CAP patients, 40 died within 28 days and out of 296 sepsis patients, 74 died within 28 days. Thus, we divided CAP and sepsis patients into survived and deceased groups, considering these two states as an ordered factor (ordinal) variable (survived < deceased).

GSE21802 Illumina human-6 v2.0 expression bead-chip whole blood gene expression profiles were used for the analysis of H1N1 infected patients [28]. The cohort consisted of 4 healthy individuals and 19 H1N1 patients (8 in early and 11 in late phase of the disease). The early phase was defined as early, from the onset of symptoms - day 0 to day 8 , and late – from day 9 and above. The statistics of the cohorts is given in Table 1 of [28], however neither sex nor age assignments were available for the patients from the GSE21802 series annotation.

GSE65682 microarrays signal intensities were pre-processed (background corrected and RMA-normalized) with the Bioconductor *oligo* package [46]. Lowly-expressed and outlier genes were identified in high dimensions using the spatial signs (*sign2*) algorithm of *mvouliter* R package with a critical value for outlier detection at 0.9.The robust principal components explained a variance of 0.95 [47]. GSE21802 signal intensities significantly above the background were quantile normalized [48]. Genes were annotated with Bioconductor *hgu219.db* and *illuminaHumanv2.db* database packages for GSE65682 (8826 genes) and GSE21802 (7240 genes) respectively.

Statistical analysis of gene expression variability and ensemble gene noise

Statistical analysis was done using R and R/Bioconductor packages [49].

To estimate the inter-individual gene expression variability for healthy, CAP and sepsis patients we accounted for age as a random effect. To this end, we used Generalized Additive Model for Location, Scale and Shape (GAMLSS) [20, 29]. In brief, for normally distributed

log-transformed microarray intensities ($Y = log(X)$, $Y \sim N(\mu_Y, \sigma_Y)$), GAMLSS allows for the modelling of both parameters of gene expression (mean and dispersion):

$$\mu_Y \sim D\beta_\mu + Zu_\mu,$$

$$log(\sigma_Y) \sim D\beta_\sigma + Zu_\sigma,$$

where $\mu_Y = \left(\mu_{Y_1}, ..., \mu_{Y_n}\right)^T$ and $\sigma_Y = \left(\sigma_{Y_1}, ..., \sigma_{Y_n}\right)^T$ are the vectors of means and dispersions for $Y = (Y_1, ..., Y_n)$. D - $n \times p$ fixed effect design matrix for the disease state (healthy, survived, deceased). $\beta_\mu = \left(\beta_{\mu_1}, ..., \beta_{\mu_r}\right)^T$ and $\beta_\sigma = \left(\beta_{\sigma_1}, ..., \beta_{\sigma_r}\right)^T$ – estimated fixed effect coefficients for mean and dispersion. Z- $n \times k$ random effect design matrix for age (age was binned into 10 deciles). $u_\mu = \left(u_{\mu_1}, ..., u_{\mu_k}\right)^T$ and $u_\sigma = \left(u_{\sigma_1}, ..., u_{\sigma_k}\right)^T$ – estimated random effect coefficients for mean and dispersion, where $u \sim N(0, \delta)$. With GAMLSS it is also straightforward to test for the significance of a factor effect on either the mean, the variance, or both with likelihood ratio test [20].

Gene ensemble lists were generated by the mapping of genes to the KEGG-annotated biological pathways or CORUM-annotated subunits of mammalian protein complexes [22, 23]. Their gene noise was estimated for each individual by calculating the variances of log-transformed expressions of genes for each ensemble (Figure 1S). Estimates of gene ensembles noise were correlated with the disease states (healthy < early phase < late phase for H1N1 and healthy < survived < deceased for CAP and sepsis) by Kendall rank correlation, treating the disease state as an ordinal variable. Linear trends between disease states and ensembles gene noise were estimated by rank-based regression [50].

Gradient boosted regression tree models

To predict the mortality of CAP and sepsis patients we trained gradient boosted regression tree models with a scalable tree boosting system XGBoost [36] using mortality within 28 days as a binary response variable, and ensemble gene noise and age as independent model features. To this end, we split individuals into discovery and validation cohorts following exactly the same partitioning as annotated in GSE65682 [8]. Then, we trained 3 models: 1) a model predicting mortality for CAP and sepsis patients, 2) a model predicting mortality for CAP patients, and 3) a model predicting mortality for sepsis patients. Models features were preselected using discovery cohorts by t test comparing ensembles gene noise for survived and deceased patients

to maximize the accuracy of XGBoost training on the discovery data sets. For CAP and sepsis (1), and sepsis (3) models, the cut-off for the model features was set at $p \leq 0.01$, and for the CAP model (2) – at $p \leq 0.05$. The XGBoost hyper tuning parameters (learning rate (η), complexity (γ), depth, *etc.*) were optimized by cross validation. To avoid overfitting, we found early epoch stopping parameters by randomly splitting of the discovery cohort into two equal folds: training and test. Then, the validation cohorts, which were hidden from feature selection and model training, were used to verify the accuracy of the final models.
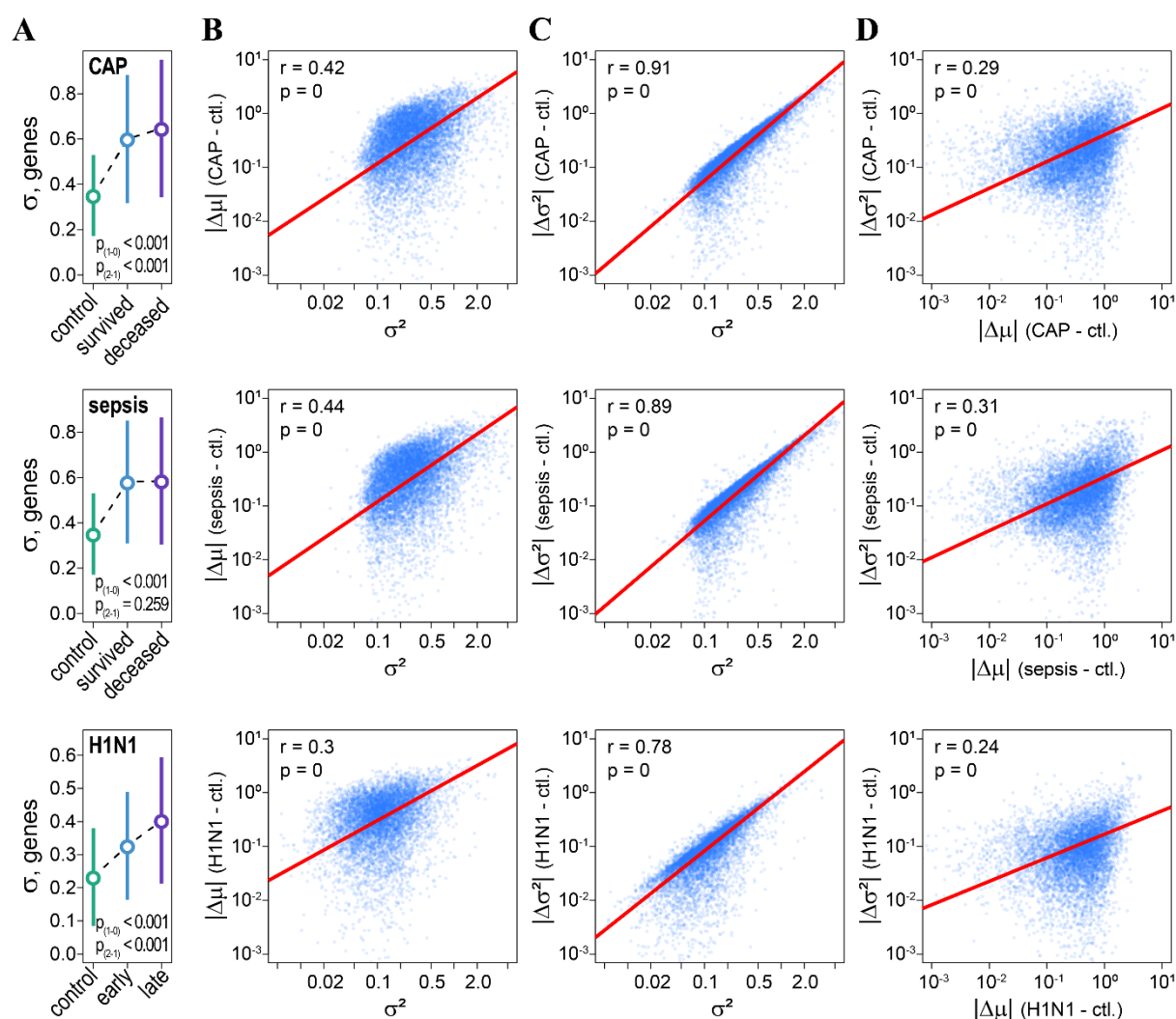
Acknowledgements

Figure 1. H1N1, CAP and sepsis coordinately affect mean gene expression and inter-individual gene expression variability.

A) Inter-individual variability in whole blood gene expression (σ) increases in CAP (top), sepsis (mid) and H1N1 (bottom) patients as compared to healthy individuals. $p_{(1-0)} - p$-values of t tests comparing differences in inter-individual gene expression variability of healthy individuals (control) with survived (CAP, sepsis) and early H1N1 infected patients. $p_{(2-1)} - p$-values of t tests comparing differences of survived (CAP, sepsis) and early H1N1 infected patients with deceased (CAP, sepsis) and late H1N1 infected patients. Circles and whiskers indicate means and standard deviations respectively.

17

B) Correlations between variances in whole blood gene expression (σ^2) and absolute changes in mean gene expression ($|\Delta\mu|$) for healthy individuals (ctl.) and patients (CAP, sepsis, H1N1). Due to the fluctuation-response relationship a magnitude of mean gene expression response depends on its variance. We estimated common variances for genes in healthy and CAP patients (top), healthy and sepsis patients (mid) and healthy and H1N1 patients (bottom).

C) Correlations between variances in whole blood gene expression (σ^2) and absolute changes in inter-individual gene expression variability ($|\Delta\sigma^2|$) for control individuals (ctl.) and patients (CAP, sepsis, H1N1).

D) Correlations between absolute changes in mean gene expression ($|\Delta\mu|$) and in inter-individual gene expression variability ($|\Delta\sigma^2|$).
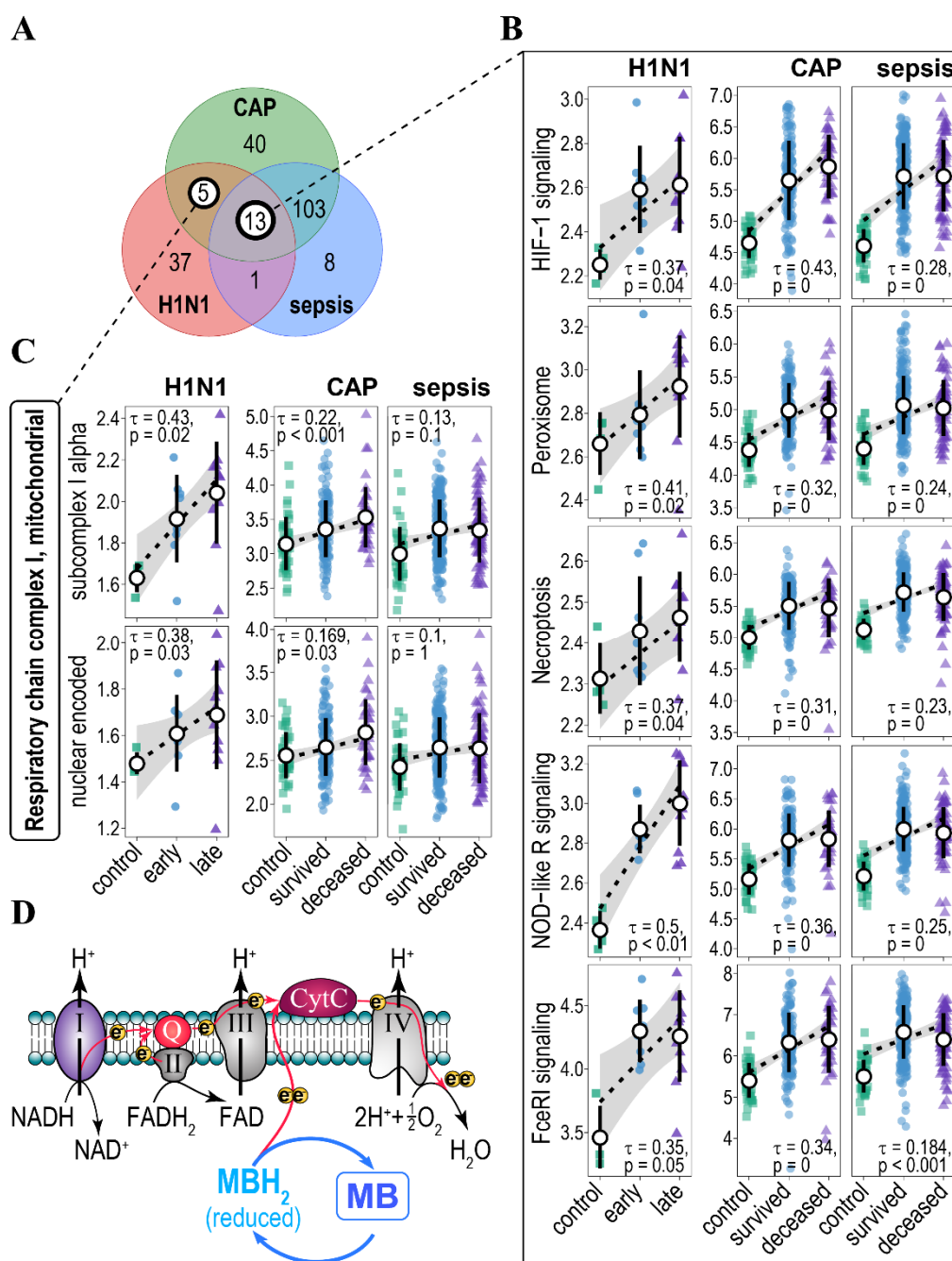
Figure 2. Association of ensemble gene noise with H1N1, CAP and sepsis disease states.

A) Venn diagram of KEGG- and CORUM-annotated biological pathways/protein complexes for which ensemble gene noise associates positively (increases) and significantly with disease state.

B) Plots of ensemble gene noise for genes involved in HIF-1 signalling, peroxisome, necroptosis, NOD-like receptor and Fc epsilon RI signalling pathways. Pathways were annotated by KEGG. Kendall tau, and FDR- (H1N1 patients) and Bonferroni- (CAP and sepsis

patients) adjusted p-values are indicated. Rank-based regression trend lines and 95% confidence bands of ensemble gene noise association with the state of disease are shown. Black circles and whiskers indicate means and standard deviations.

C) Plots of ensemble gene noise for genes encoding CORUM-annotated subunits of mitochondrial respiratory chain complex I (subcomplex I alpha – top panel and nuclear encoded subunits – bottom panel). Rank-based regression trend lines and 95% confidence bands of ensemble gene noise association with the state of disease are shown.

D) Methylene Blue (MB) acts as an alternative electron donor to the electron transport chain (red arrows) by shuttling between redox states (MB – MBH_2) and, thus, bypassing respiratory chain complex I. Respiratory chain complex I-IV and their substrates are indicated, Q – coenzyme Q10, CytC – cytochrome C. Electrons are indicated as yellow circles.
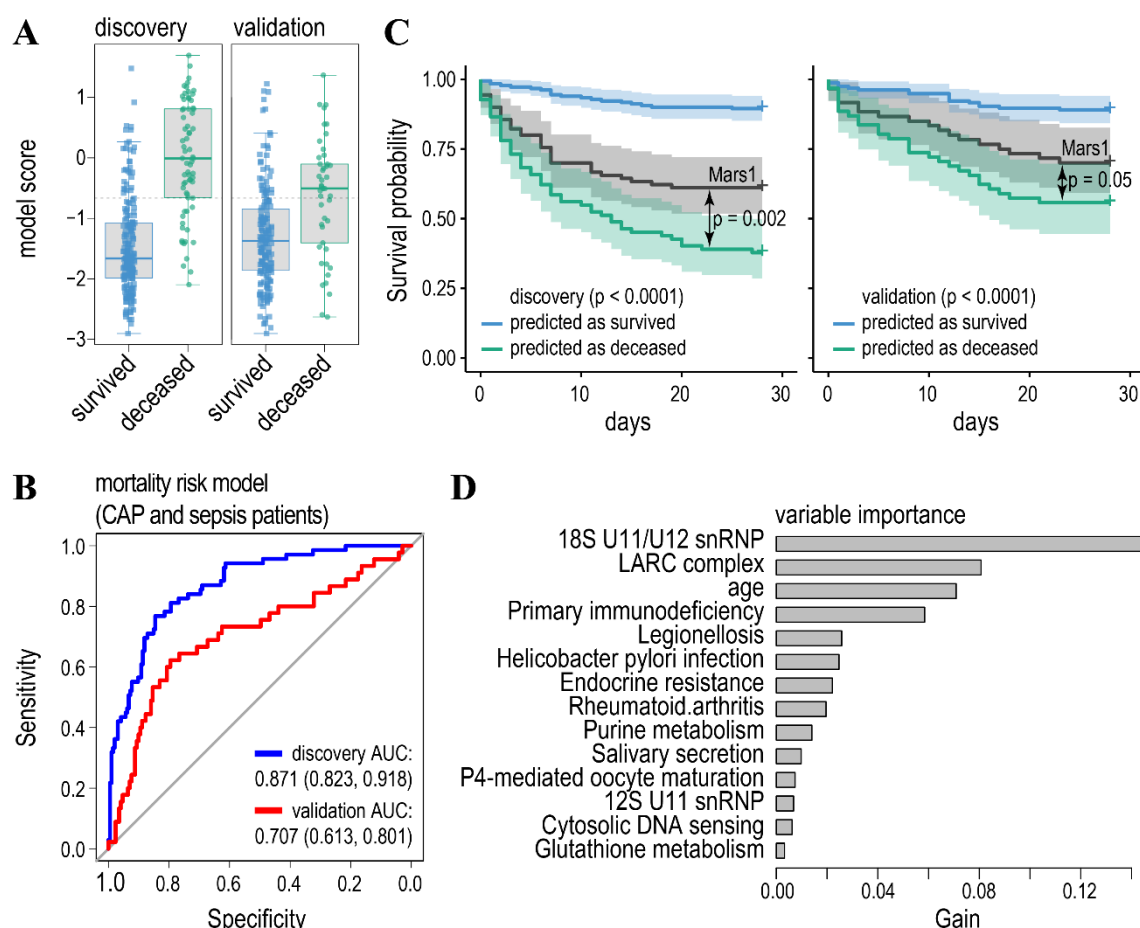
Figure 3. Model predicting mortality/survival of CAP and sepsis patients.

A) Boxplots of the model scores predicting mortality/survivorship in the discovery (left) and validation (right) cohorts. The model was trained on the same as published discovery cohort by the gradient boosted regression tree and validated on independent cohort[8]. Dashed lines indicate threshold levels of classification. The threshold was calculated by maximizing a product of the specificity and sensitivity of the model prediction in the discovery cohort. Further details of model accuracy are given in Tables 2 and S2.

B) Receiver operating characteristic curves (ROC) for the model predicting mortality (end point – survival or death within 28 days after treatment) in CAP and sepsis patients (blue line – discovery cohort, red line – validation cohort). Features were selected by the t-test comparing ensemble gene noise between the survived and deceased patients in the discovery cohort to achieve maximum prediction accuracy for the discovery cohort. Values for the area under the ROC curve (AUC) are indicated.

21

C) Survival probability for the patients predicted to have low (blue line) and high (green line) risk of mortality for the discovery (left panel) and validation (right panel) cohorts. p-values indicate significant differences in hazards for the predicted classes (survival/mortality) according to the Cox proportional-hazards model. Black lines - survival probability of patients with Mars1 endotype [8] was compared with the predicted deceased class for the discovery and validation cohorts.

D) Variable importance of the model ranks ensemble gene noise features according to their relative contribution (gain).
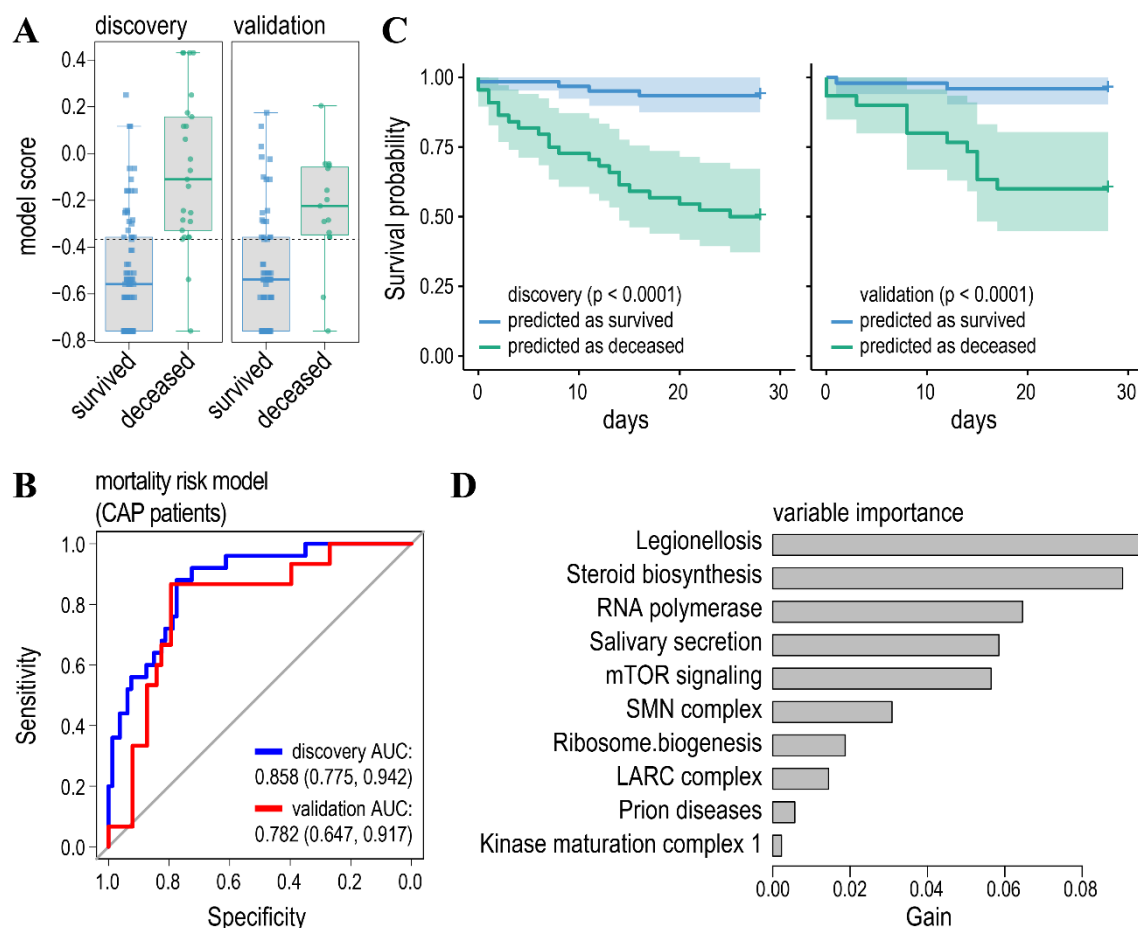
Figure 4. Model predicting mortality/survival of CAP patients. A) Boxplots of the model scores predicting mortality/survivorship in the discovery (left) and validation (right) cohorts. **B)** ROC curves for the model predicting mortality in CAP patients in the discovery (blue line) and validation (red line) cohorts. Cohorts were partitioned as in [8]. **C)** Survival probability for the patients predicted to have low (blue line) and high (green line) risk of mortality for the discovery (left panel) and validation (right panel) cohorts. **D)** Relative contribution of ensemble gene noise features to the model.
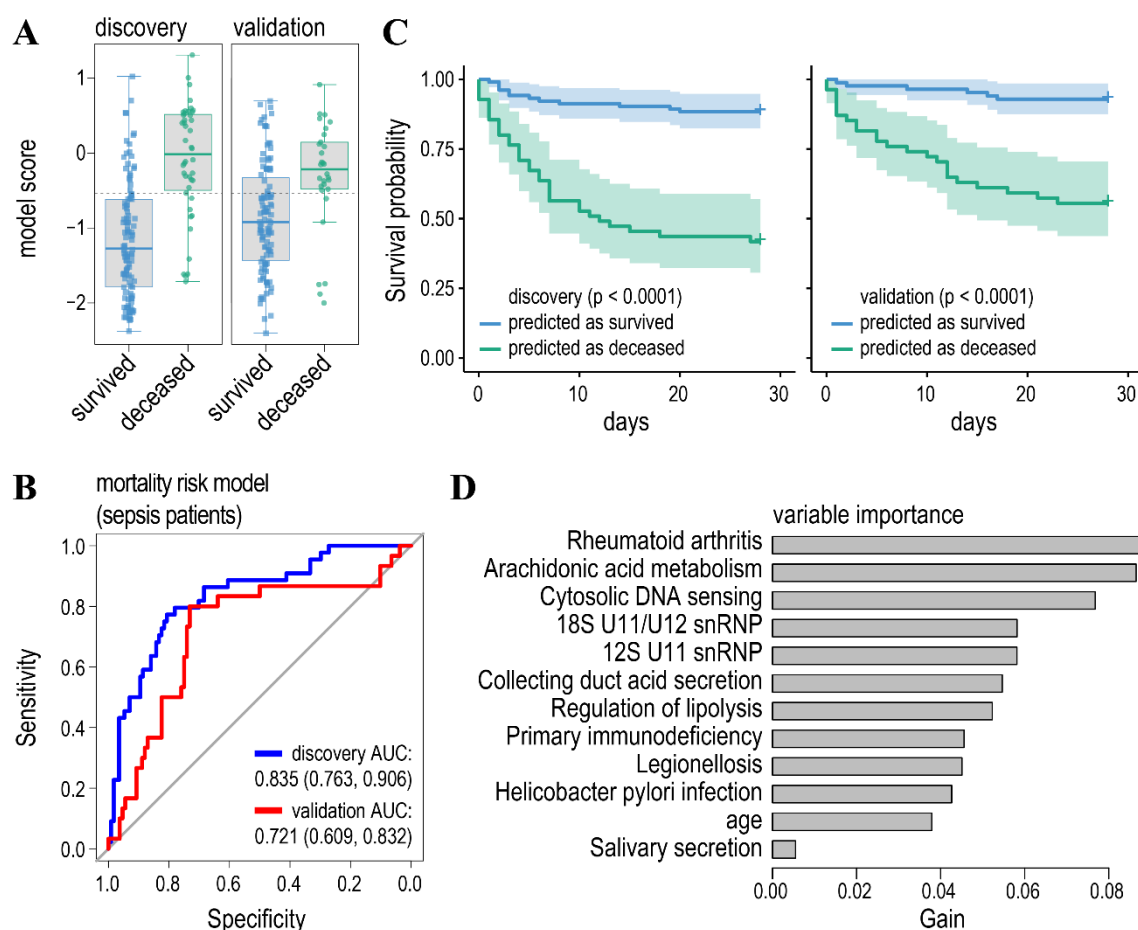
Figure 5. Model predicting mortality/survival of sepsis patients. A) Boxplots of the model scores. **B)** ROC curves for the model predicting mortality in sepsis patients. **C)** Survival probability for the patients predicted to have low (blue line) and high (green line) risk of mortality. **D)** Relative contribution of ensemble gene noise features to the model.

Table 1. Role in sepsis of the pathways for which ensemble gene noise associates positively (increases) with the disease states (healthy < early/survived < late/deceased)

Pathways/Complexes	Role in sepsis pathology	Reference
KEGG: HIF-1 signalling pathway	Metabolic reprogramming of innate immune cells during the hyperinflammatory and immunotolerant phases of sepsis.	[51, 52]
KEGG: Peroxisome	Defective peroxisome recycling alters cellular redox homeostasis and leads to exaggerated oxidative stress response to endotoxin (infection) and sepsis.	[43]
KEGG: Necroptosis	Necroptosis is implicated in pulmonary diseases and sepsis-associated organ injury.	[53, 54]
KEGG: NOD-like receptor signalling pathway	Activation of Toll-like and NOD-like receptor signalling protects mice from polymicrobial sepsis-associated lethality.	[55]
KEGG: Fc epsilon RI signalling pathway	Fc receptors bind to antibodies attached to invading pathogens and their up-regulation can serve a potential biomarker for sepsis. Mice deficient for FCER1G gene encoding the γ-subunit of Fc epsilon RI show increased resistance to sepsis.	[56, 57]
KEGG: Autophagy - other	Autophagy is an adaptive protective process that eliminates damaged proteins, organelles and pathogens. It is thought to be a promising target in treatment of sepsis.	[58]
KEGG: Biosynthesis of amino acids	Sepsis results in significant disorders in amino acids metabolism.	[59]
KEGG: Glucagon signalling pathway	Glucagon levels negatively associate with clinical outcome in sepsis patients	[60]
KEGG: Propanoate (propionate) metabolism	Propionic acidaemia caused by altered propionate metabolism often results in sepsis and death.	[61]
KEGG: Circadian rhythm	There is accumulating evidence for association between circadian misalignment and severity of inflammatory responses in sepsis.	[62]
KEGG: Dopaminergic synapse	Dopamine mediates neuroimmune communications and dopaminergic is implicated in inflammation and sepsis.	[63, 64]
KEGG: Amyotrophic lateral sclerosis (ALS)	ALS patients often develop pulmonary insufficiency and have increased risk of sepsis.	[65]
CORUM: Respiratory chain complex I, mitochondrial	Mitochondrial disfunction resulting in reduced respiratory chain complex I activity and low ATP levels is a whole mark for sepsis.	[30]
KEGG: Osteoclast differentiation	Mean expression of osteoclast differentiation genes is up-regulated in human septic shock.	[66]
KEGG: Tight junction	Sepsis disrupts intestinal barrier which leads to a multiple organ dysfunction syndrome and alters the expression of tight junction proteins.	[67]

Table 2. Prediction accuracy of the models

Metric	CAP/sepsis patients		CAP patients		Sepsis patients	
	discovery	validation	discovery	validation	discovery	validation
bACC	0.799	0.701	0.802	0.798	0.779	0.761
Sensitivity	0.754	0.6	0.88	0.867	0.75	0.8
Specificity	0.845	0.801	0.725	0.73	0.807	0.722

Supplementary material

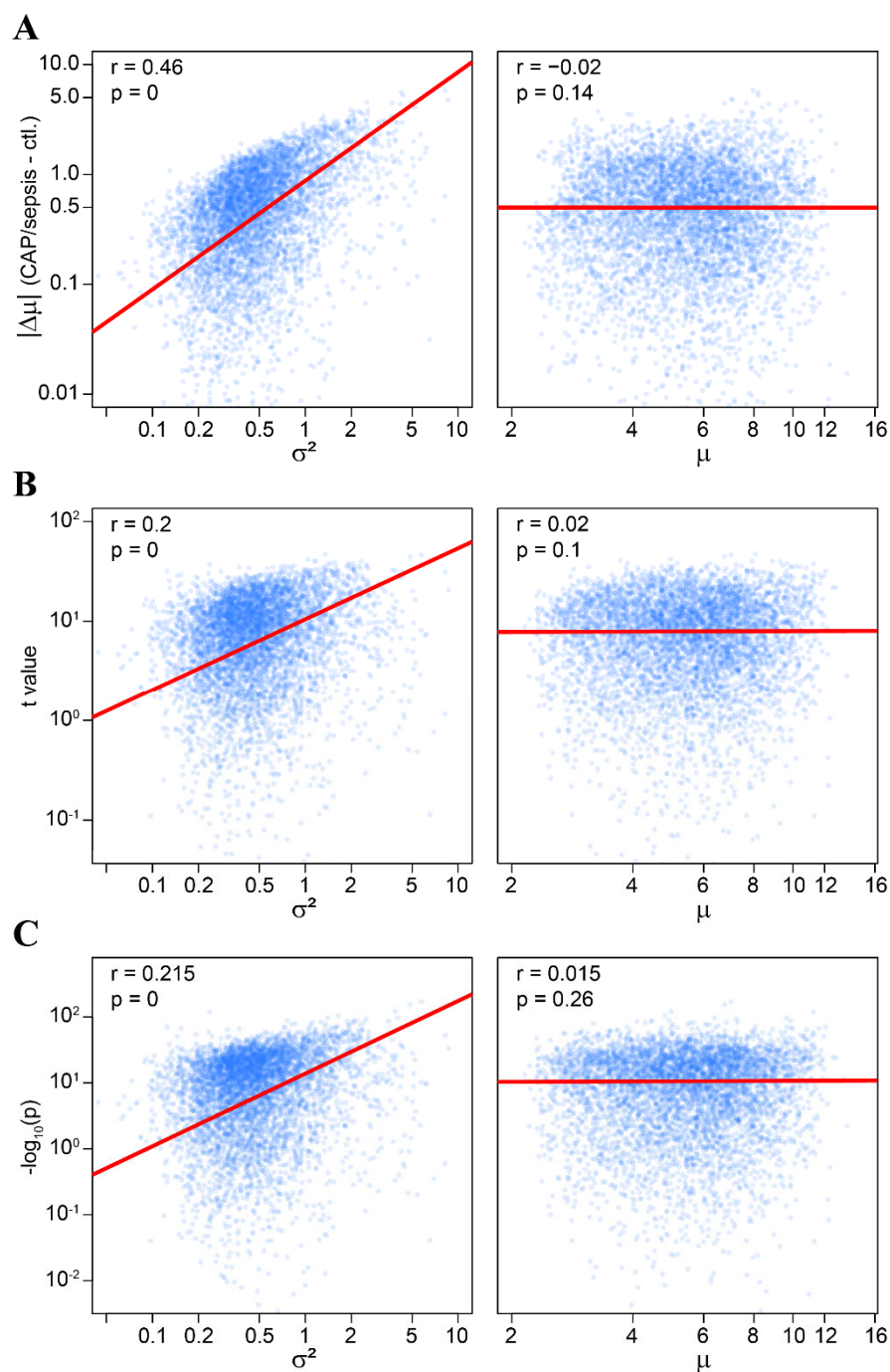**Figure S1. Fluctuation-response relation biases statistical inference of DGE. A)** The difference of means of log-transformed bell-shaped gene expression values ($Y = log(X)$) are proportional to the variance or the squared coefficient of variation (cv^2) of untransformed variable (X): $|\bar{Y}_1 - \bar{Y}_0| \sim \sigma_Y^2 \approx \frac{\sigma_X^2}{\bar{X}^2} = cv_X^2$. Considering RNA copy number (X) to be mixed-

Poisson (negative-binomial as a specific case) random variable, for large X, cv_X^2 approximates overdispersion parameter or biological coefficient of variation (bcv^2) [20]. A scatterplot on the left panel illustrates fluctuation-response relation (a correlation between absolute differences and variances of log-transformed expression values) for whole-blood gene expression profiles of healthy (ctl.) individuals and CAP pneumonia/sepsis patients. The data has been taken from [8]. This relation is monotonic, but non-linear, suggesting a deviation from linear coupling between fluctuation (pneumonia/sepsis) and gene expression response. There is no correlation between differences and means of log-transformed expression values (right panel). **B-C)** In the presence of fluctuation-response relation statistical inference will be biased. For example, Student's t-test often used to assess DGE will relate positively to a variance of log-transformed gene expression as: $t \sim \frac{|\bar{Y}_1 - \bar{Y}_0|}{\sigma_Y} \sim \frac{\sigma_Y^2}{\sigma_Y} = \sigma_Y$. A scatterplot on the left panel shows correlation between Student's t statistic and variance (**B**) and Bonferroni-adjusted p values and variance (**C**). There is no correlation of t and p values with means of log-transformed expression values (right panel).
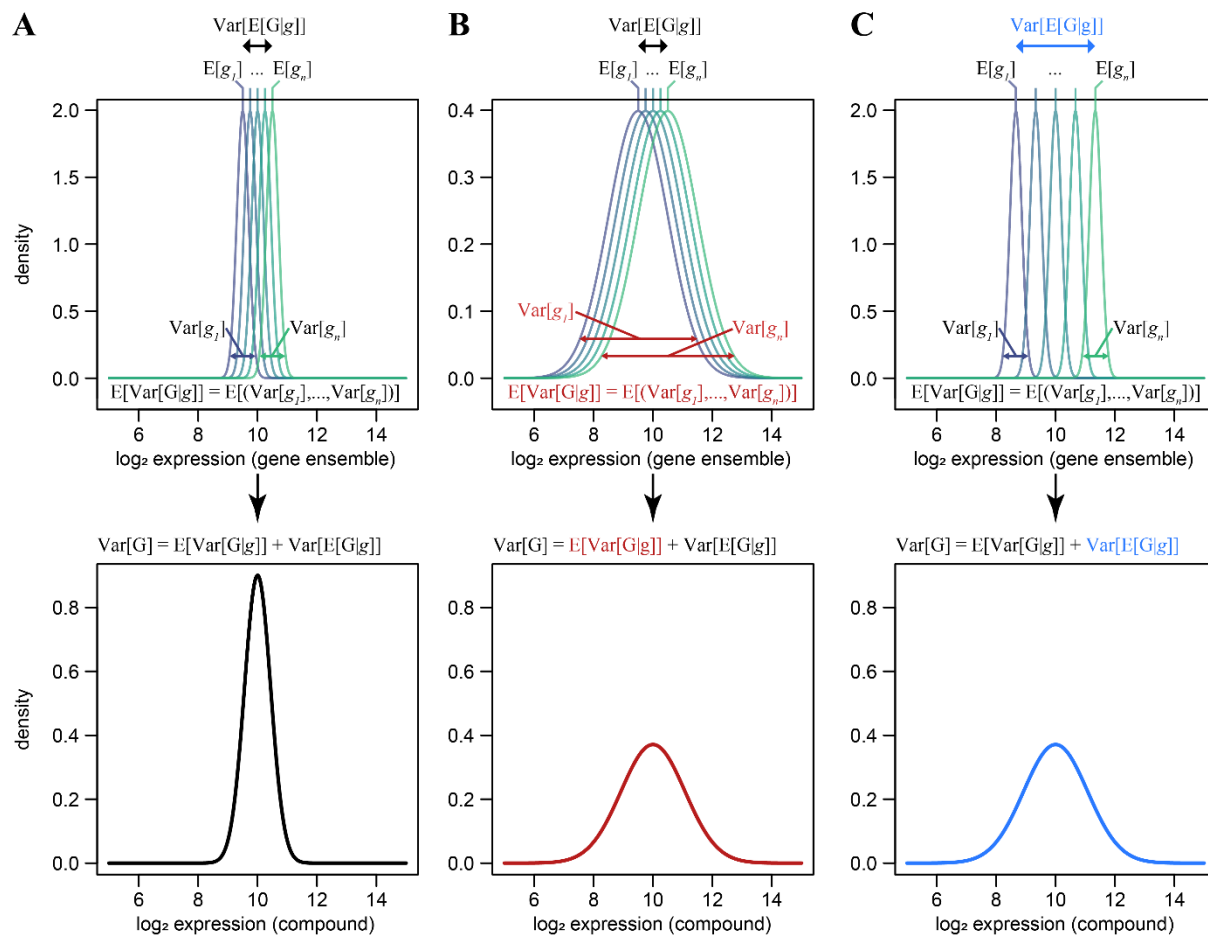
Figure S2. Ensemble gene noise. A) Let G be a collection of genes $(g_1, \dots, g_n)$ belonging to either a biological pathway or encoding subunits of a protein complex. Then, from the law of total variance $\mathrm{Var}[G] = \mathrm{E}\big[\mathrm{Var}[G|g]\big] + \mathrm{Var}\big[\mathrm{E}[G|g]\big]$, *i.e.* ensemble gene noise $(\mathrm{Var}[G])$ sums from the expected value of genes' variances $(\mathrm{E}\big[\mathrm{Var}[G|g]\big])$ and the variance in genes' mean expression $(\mathrm{Var}\big[\mathrm{E}[G|g]\big])$. The top panel illustrates hypothetical distributions of expressions of genes in ensemble (g_i), the bottom panel is derived distribution of gene ensemble $(G = (g_1, \dots, g_n))$. **B, C)** Top panel, changes in variances (**B**) and/or expectations (**C**) of genes expression will eventually change ensemble gene noise (bottom panel).
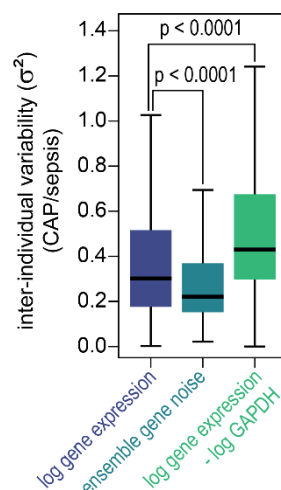
Figure S3. Comparison of inter-individual variability for log gene expression and ensemble gene noise for CAP and sepsis patients. Boxplots illustrating population variances for log gene expression, ensemble gene noise and log gene expression normalized to the GAPDH. Inter-individual variability is significantly less for ensemble gene noise as compared to the log gene expression (according to t-test) and it is higher for GAPDH normalized log gene expressions. The latter follows from the fact that $\text{Var}[log(X) - log(\text{GAPDH})] \approx \text{Var}[log(X)] + \text{Var}[log(\text{GAPDH})]$. Thus, estimating DGE by PCR, which usually requires normalization to some housekeeping gene, results in increased inter-individual variability. Ensemble gene noise can be estimated from PCR without normalization to a reference gene.

Figure S4. Fluctuation-response relation for ensemble gene noise. Differences of means of ensembles gene noise are proportional to the inter-individual variability in ensembles gene noise. A scatterplot on the left panel illustrates fluctuation-response relation (a correlation between absolute differences and variances of ensembles gene noise) for whole-blood gene expression profiles of healthy (ctl.) individuals and CAP pneumonia/sepsis patients. There is also a modest correlation between absolute changes in ensembles gene noise and the means (right panel).

Table S1. Kendall correlation of ensemble gene noise with H1N1, CAP and sepsis disease states.

A) H1N1 (p ≤ 0.05) ∩ CAP (p ≤ 0.05) ∩ sepsis (p ≤ 0.05)

Gene ensemble (KEGG/CORUM)	H1N1 (τ, p)	CAP (τ, p)	Sepsis (τ, p)	genes
KEGG: HIF-1 signalling pathway	τ = 0.373 p = 0.0357	τ = 0.429 p = 0	τ = 0.283 p = 0	AKT1; AKT2; ALDOA; ARNT; CAMK2D; CAMK2G; CDKN1A; CDKN1B; CREBBP; CUL2; CYBB; EDN1; EGF; EGLN1; EGLN2; EIF4E2; EIF4EBP1; ELOB; ELOC; ENO1; ENO2; EP300; GAPDH; HIF1A; HK1; HK2; HK3; HMOX1; IFNG; IFNGR1; IFNGR2; IGF1R; IL6R; INSR; LDHA; LTBR; MAP2K1; MAP2K2; MAPK3; MKNK1; MKNK2; NFKB1; NPPA; PDHA1; PDHB; PDK1; PFKFB3; PFKL; PGK1; PIK3CA; PIK3CB; PIK3CD; PIK3R1; PLCG2; PRKCB; RBX1; RPS6; RPS6KB1; RPS6KB2; SLC2A1; STAT3; TFRC; TIMP1; TLR4; VHL
KEGG: Peroxisome	τ = 0.412 p = 0.0219	τ = 0.323 p = 0	τ = 0.235 p = 0	ABCD1; ABCD3; ACAA1; ACOT8; ACOX1; ACOX2; ACOX3; ACSL1; ACSL3; ACSL4; ACSL5; ACSL6; AGPS; AMACR; CAT; CRAT; DHRS4; DHRS4L1; ECI2; FAR1; FAR2; GNPAT; GSTK1; HACL1; HMGCL; HSD17B4; IDH1; IDH2; MPV17; MVK; NUDT19; NUDT7; PECR; PEX11B; PEX11G; PEX16; PEX5; PEX6; PEX7; PRDX1; PRDX5; PXMP2; PXMP4; SLC27A2; SOD1; SOD2
KEGG: Necroptosis	τ = 0.373 p = 0.0357	τ = 0.309 p = 0	τ = 0.229 p = 0	AIFM1; BAX; BID; BIRC2; BIRC3; CAMK2D; CAMK2G; CAPN1; CAPN2; CASP1; CASP8; CFLAR; CHMP1A; CHMP1B; CHMP2A; CHMP2B; CHMP3; CHMP4A; CHMP4B; CHMP5; CHMP6; CHMP7; CYBB; CYLD; DNM1L; EIF2AK2; FADD; FAF1; FAS; FASLG; FTH1; FTL; GLUD1; GLUD2; GLUL; H2AFX; H2AFY; H2AFZ; HIST1H2AA; HIST1H2AC; HIST1H2AD; HIST1H2AE; HIST2H2AA3; HIST2H2AA4; HIST2H2AC; HIST3H2A; HMGB1; HSP90AA1; HSP90AB1; IFNAR1; IFNAR2; IFNG; IFNGR1; IFNGR2; IL1A; IL1B; IRF9; JAK1; JAK2; JAK3; JMJD7-PLA2G4B; MLKL; NLRP3; PARP4; PLA2G4A; PLA2G4F; PPID; PYCARD; PYGB; PYGL; RBCK1; RIPK1; RIPK3; RNF31; SHARPIN; SLC25A5; SLC25A6; SMPD1; SPATA2L; SQSTM1; STAT1; STAT2; STAT3; STAT4; STAT5A; STAT5B; STAT6; TICAM1; TLR4; TNF; TNFAIP3; TNFRSF10A; TNFRSF10B; TNFRSF1A; TNFSF10; TRADD; TRAF5; TRPM7; TYK2; USP21; VDAC1; VDAC2; VDAC3; VPS4A; VPS4B; XIAP; ZBP1
KEGG: NOD-like receptor signalling pathway	τ = 0.501 p = 0.0062	τ = 0.357 p = 0	τ = 0.252 p = 0	AIM2; ATG12; ATG5; BCL2L1; BIRC2; BIRC3; BRCC3; CAMP; CARD16; CARD17; CARD6; CARD8; CARD9; CASP1; CASP4; CASP5; CASP8; CCL2; CCL5; CHUK; CTSB; CXCL1; CXCL8; CYBA; CYBB; DEFA1B; DEFA3; DEFA4; DHX33; DNM1L; ERBIN; FADD; GABARAP; GABARAPL1; GABARAPL2; GBP1; GBP2; GBP3; GBP4; GBP5; GSDMD; HSP90AA1; HSP90AB1; IFI16; IFNAR1; IFNAR2; IKBKB; IKBKE; IKBKG; IL18; IL1B; IRAK4; IRF3; IRF7; IRF9; JAK1; JUN; MAP3K7; MAPK13; MAPK14; MAPK3; MAVS; MCU; MEFV; MFN1; MFN2; MYD88; NAIP; NAMPT; NEK7; NFKB1; NFKBIA; NLRC4; NLRP1; NLRP12; NLRP3; NLRP7; NLRX1; NOD2; OAS1; OAS2; OAS3; P2RX7; PLCB1; PLCB2; PRKCD; PSTPIP1; PYCARD; RBCK1; RHOA; RIPK1; RIPK2; RIPK3; RNASEL; RNF31; SHARPIN; STAT1; STAT2; SUGT1; TAB1; TAB2; TAB3; TANK; TBK1; TICAM1; TLR4; TMEM173; TNF; TNFAIP3; TRAF3; TRAF5; TRAF6; TRIP6; TRPM2; TRPM7; TXN; TXN2; TXNIP; TYK2; VDAC1; VDAC2; VDAC3; XIAP

Pathway				Genes
KEGG: Fc epsilon RI signalling pathway	τ = 0.353 p = 0.0493	τ = 0.342 p = 0	τ = 0.184 p = 0.00007	AKT1; AKT2; ALOX5; ALOX5AP; BTK; FCER1A; FCER1G; FYN; GAB2; GRB2; INPP5D; JMJD7-PLA2G4B; KRAS; LAT; LCP2; LYN; MAP2K1; MAP2K2; MAP2K3; MAP2K4; MAP2K6; MAP2K7; MAPK13; MAPK14; MAPK3; MS4A2; NRAS; PDPK1; PIK3CA; PIK3CB; PIK3CD; PIK3R1; PLA2G4A; PLA2G4F; PLCG2; RAC2; RAF1; SOS1; SOS2; SYK; TNF; VAV1
KEGG: Autophagy - other	τ = 0.491 p = 0.0067	τ = 0.199 p = 0.0015	τ = 0.137 p = 0.0302	ATG101; ATG12; ATG13; ATG2A; ATG2B; ATG3; ATG4A; ATG4C; ATG5; ATG7; ATG9A; BECN1; GABARAP; GABARAPL1; GABARAPL2; IGBP1; MLST8; PIK3C3; PIK3R4; PPP2CA; PPP2CB; WIPI1; WIPI2
KEGG: Biosynthesis of amino acids	τ = 0.363 p = 0.0455	τ = 0.336 p = 0	τ = 0.274 p = 0	ACO2; ALDH18A1; ALDOA; ALDOC; ARG1; ARG2; ASL; ASNS; BCAT1; BCAT2; CBS; CS; ENO1; ENO2; GAPDH; GLUL; GOT1; GOT2; IDH1; IDH2; IDH3B; IDH3G; MAT2A; MAT2B; MTR; PFKL; PFKM; PFKP; PGAM1; PGK1; PHGDH; PKM; PRPS1; PRPS2; PSPH; PYCR2; PYCR3; RPIA; SDSL; SHMT1; SHMT2; TALDO1; TKT; TKTL1; TPI1
CORUM: TRAPP complex	τ = 0.412 p = 0.0219	τ = 0.276 p = 0	τ = 0.144 p = 0.0136	TRAPPC1; TRAPPC10; TRAPPC11; TRAPPC12; TRAPPC2; TRAPPC2L; TRAPPC3; TRAPPC4; TRAPPC5; TRAPPC6A; TRAPPC6B; TRAPPC8; TRAPPC9
KEGG: Glucagon signalling pathway	τ = 0.403 p = 0.0299	τ = 0.211 p = 0.0004	τ = 0.145 p = 0.0122	ACACA; AKT1; AKT2; ATF4; CALM1; CALM2; CALML4; CAMK2D; CAMK2G; CPT1A; CPT1B; CREB1; CREB3; CREB3L1; CREB5; CREBBP; CRTC2; EP300; FBP1; GNAQ; GNAS; GYS1; LDHA; LDHB; PCK2; PDE3B; PDHA1; PDHB; PFKL; PGAM1; PHKA2; PHKB; PHKG2; PKM; PLCB1; PLCB2; PPP3CA; PPP3CB; PPP3CC; PPP3R1; PPP4C; PPP4R3A; PPP4R3B; PRKAA1; PRKAB1; PRKACB; PRKAG1; PRKAG2; PRMT1; PYGB; PYGL; SIK1; SIK2; SIRT1; SLC2A1
KEGG: Propanoate (propionate) metabolism	τ = 0.393 p = 0.03	τ = 0.182 p = 0.0081	τ = 0.137 p = 0.0298	ABAT; ACACA; ACADM; ACAT1; ACAT2; ACSS1; ACSS2; ALDH6A1; BCKDHA; BCKDHB; DLD; ECHDC1; ECHS1; HADHA; HIBCH; LDHA; LDHB; PCCA; PCCB; SUCLA2; SUCLG1
KEGG: Circadian rhythm	τ = 0.393 p = 0.03	τ = 0.353 p = 0	τ = 0.288 p = 0	ARNTL; BHLHE40; BTRC; CLOCK; CREB1; CRY1; CSNK1D; CUL1; FBXL3; FBXW11; NR1D1; PER1; PRKAA1; PRKAB1; PRKAG1; PRKAG2; RBX1; RORA; SKP1
KEGG: Dopaminergic synapse	τ = 0.599; 0.001005	τ = 0.355; 0	τ = 0.252; 0	AKT1; AKT2; ARNTL; ARRB2; ATF4; ATF6B; CACNA1A; CALM1; CALM2; CALML4; CAMK2D; CAMK2G; CLOCK; CREB1; CREB3; CREB3L1; CREB5; FOS; GNAI2; GNAI3; GNAQ; GNAS; GNB1; GNB2; GNB4; GNB5; GNG10; GNG11; GNG2; GNG7; GNG8; GSK3A; GSK3B; KIF5B; MAOA; MAOB; MAPK13; MAPK14; PLCB1; PLCB2; PPP1CA; PPP1CB; PPP1CC; PPP2CA; PPP2CB; PPP2R2A; PPP2R2B; PPP2R2D; PPP2R3B; PPP2R3C; PPP2R5A; PPP2R5B; PPP2R5C; PPP2R5D; PPP2R5E; PPP3CA; PPP3CB; PPP3CC; PRKACB; PRKCB
KEGG: Amyotrophic lateral sclerosis (ALS)	τ = 0.461 p = 0.0102	τ = 0.292 p = 0	τ = 0.189 p = 0.00003	ALS2; BAD; BAX; BCL2L1; BID; CASP1; CASP3; CASP9; CAT; CCS; DAXX; DERL1; GPX1; GRIN1; MAP2K3; MAP2K6; MAP3K5; MAPK13; MAPK14; PPP3CA; PPP3CB; PPP3CC; PPP3R1; RAB5A; SOD1; TNF; TNFRSF1A; TNFRSF1B; TOMM40; TOMM40L

B) H1N1 (p ≤ 0.05) ∩ CAP (p ≤ 0.05) ∩ sepsis (p > 0.05)

Gene ensemble (KEGG/CORUM)	H1N1 (τ, p)	CAP (τ, p)	Sepsis (τ, p)	genes
CORUM: Respiratory chain complex I (subcomplex I alpha), mitochondrial	τ = 0.432 p = 0.0178	τ = 0.222 p = 0.0001	τ = 0.127 p = 0.0929	NDUFA1; NDUFA10; NDUFA11; NDUFA12; NDUFA13; NDUFA2; NDUFA3; NDUFA5; NDUFA6; NDUFA7; NDUFA8; NDUFA9; NDUFAB1; NDUFB4; NDUFB7; NDUFS1; NDUFS2; NDUFS3; NDUFS6; NDUFS7; NDUFS8; NDUFV1; NDUFV2
CORUM: Respiratory chain complex I (nuclear encoded subunits), mitochondrial	τ = 0.383 p = 0.0323	τ = 0.169 p = 0.0271	τ = 0.092 p = 1	NDUFA1; NDUFA10; NDUFA11; NDUFA12; NDUFA13; NDUFA2; NDUFA3; NDUFA4; NDUFA5; NDUFA6; NDUFA7; NDUFA8; NDUFA9; NDUFAB1; NDUFB1; NDUFB10; NDUFB11; NDUFB2; NDUFB3; NDUFB4; NDUFB5; NDUFB6; NDUFB7; NDUFB8; NDUFB9; NDUFC1; NDUFC2; NDUFS1; NDUFS2; NDUFS3; NDUFS4; NDUFS5; NDUFS6; NDUFS7; NDUFS8; NDUFV1; NDUFV2
KEGG: Osteoclast differentiation	τ = 0.373 p = 0.0357	τ = 0.279 p = 0	τ = 0.13 p = 0.066	ACP5; AKT1; AKT2; BLNK; BTK; CHUK; CREB1; CSF1R; CTSK; CYBA; CYLD; FCGR1A; FCGR2A; FCGR2B; FCGR2C; FCGR3A; FCGR3B; FHL2; FOS; FOSL2; FYN; GAB2; GRB2; IFNAR1; IFNAR2; IFNG; IFNGR1; IFNGR2; IKBKB; IKBKG; IL1A; IL1B; IL1R1; IRF9; ITGB3; JAK1; JUN; JUNB; JUND; LCK; LCP2; LILRA1; LILRA2; LILRA4; LILRA5; LILRA6; LILRB1; LILRB2; LILRB3; LILRB4; MAP2K1; MAP2K6; MAP2K7; MAP3K14; MAP3K7; MAPK13; MAPK14; MAPK3; MITF; NCF1; NCF2; NCF4; NFATC1; NFKB1; NFKB2; NFKBIA; OSCAR; PIK3CA; PIK3CB; PIK3CD; PIK3R1; PLCG2; PPARG; PPP3CA; PPP3CB; PPP3CC; PPP3R1; RELB; SIRPA; SIRPB1; SIRPG; SOCS1; SOCS3; SPI1; SQSTM1; STAT1; STAT2; SYK; TAB1; TAB2; TGFB1; TGFBR2; TNF; TNFRSF1A; TRAF6; TYK2; TYROBP
KEGG: Tight junction	τ = 0.412 p = 0.0219	τ = 0.208 p = 0.0006	τ = 0.127 p = 0.0933	ACTB; ACTG1; ACTN1; ACTN4; ACTR2; ACTR3; ARHGAP17; ARHGEF18; ARHGEF2; CD1C; CDC42; CDK4; CLDN15; CLDN5; CLDN9; DLG1; EPB41L4B; EZR; F11R; HCLS1; HSPA4; ITGB1; JUN; LLGL2; MAP2K7; MAP3K1; MAP3K5; MICALL2; MSN; MYH9; MYL12A; MYL12B; MYL6; MYL6B; MYL9; NEDD4; NEDD4L; OCLN; PARD3; PARD6A; PCNA; PPP2CA; PPP2CB; PPP2R2A; PPP2R2B; PPP2R2D; PRKAA1; PRKAB1; PRKACB; PRKAG1; PRKAG2; PRKCE; PRKCI; PRKCZ; RAB13; RAB8A; RAB8B; RAP1A; RAP2C; RAPGEF2; RAPGEF6; RHOA; ROCK1; ROCK2; RUNX1; SCRIB; SLC9A3R1; STK11; TIAM1; TJP2; TUBA1A; TUBA1B; TUBA1C; TUBA4A; TUBA8; VASP; WASL; WHAMM; YBX3
KEGG: Axon guidance	τ = 0.461 p = 0.0102	τ = 0.165 p = 0.0423	τ = 0.074 p = 1	ABLIM1; ARHGEF12; BMPR2; CAMK2D; CAMK2G; CDC42; CDK5; CFL1; CXCR4; EFNA1; EFNB3; EPHA4; EPHB1; EPHB4; FES; FYN; GNAI2; GNAI3; GSK3B; ILK; ITGB1; KRAS; LIMK2; LRRC4; MAPK3; MYL12A; MYL12B; MYL5; MYL9; NCK1; NCK2; NFATC3; NRAS; NTNG2; PAK1; PAK2; PAK3; PAK4; PARD3; PARD6A; PDK1; PIK3CA; PIK3CB; PIK3CD; PIK3R1; PLCG2; PLXNA2; PLXNA3; PLXNB2; PLXNC1; PPP3CA; PPP3CB; PPP3CC; PPP3R1; PRKCZ; PTPN11; RAC2; RAF1; RASA1; RHOA; ROBO3; ROCK1; ROCK2; RRAS; RYK; SEMA4A; SEMA4B; SEMA4C; SRGAP2; SRGAP3; SSH1; SSH2; SSH3

Table S2. Confusion tables for the models predicting mortality/survivorship of CAP/sepsis patients

A) CAP and sepsis patients

discovery		Actual	
		survived	deceased
predicted	survived	**164 (84.5%)**	17 (24.6%)
	deceased	30 (15.5%)	**52 (75.4%)**

validation		Actual	
		survived	deceased
predicted	survived	**137 (80.1%)**	18 (40.0%)
	deceased	34 (19.9%)	**27 (60.0%)**

B) CAP patients

discovery		Actual	
		survived	deceased
predicted	survived	**58 (72.5%)**	3 (12.0%)
	deceased	22 (27.5%)	**22 (88.0%)**

validation		Actual	
		survived	deceased
predicted	survived	**46 (73.0%)**	2 (13.3%)
	deceased	17 (27.0%)	**13 (86.7%)**

C) Sepsis patients

discovery		Actual	
		survived	deceased
predicted	survived	**92 (80.7%)**	11 (25.0%)
	deceased	22 (19.3%)	**33 (75.0%)**

validation		Actual	
		survived	deceased
predicted	survived	**78 (72.2%)**	6 (20.0%)
	deceased	30 (27.8%)	**24 (80.0%)**

Table S3. Relative contribution (gain %, cover %, frequency %) of ensemble gene noise features to the models predicting mortality/survivorship of CAP/sepsis patients

Gene ensemble (KEGG/CORUM)	Gain; Cover; Frequency			genes
	CAP/sepsis	CAP	Sepsis	
CORUM: 18S U11-U12 snRNP	14.537; 10.6; 8.054		5.818; 5.966; 5.376	DHX15; PDCD7; PRPF8; RNPC3; SF3B1; SF3B2; SF3B3; SF3B4; SF3B5; SF3B6; SNRNP25; SNRNP35; SNRNP48; SNRPB; SNRPD1; SNRPD2; SNRPD3; SNRPE; SNRPF; SNRPG; YBX1; ZCRB1; ZMAT5; ZRSR2
CORUM: LARC complex (LCR-associated remodelling complex)	8.069; 7.116; 6.04	1.444; 3.508; 6.061		ACTB; ACTL6A; ARID1A; CHD4; DPF2; GATAD2B; HDAC1; MBD2; MBD3; RBBP4; SMARCA4; SMARCC2; SMARCD2; SMARCE1
KEGG: Primary immunodeficiency	5.845; 6.663; 6.711		4.555; 5.175; 4.301	ADA; BLNK; BTK; CD19; CD3D; CD3E; CD40; CD79A; CD8A; CD8B; CIITA; DCLRE1C; IKBKG; IL2RG; IL7R; JAK3; LCK; ORAI1; PTPRC; RFX5; RFXANK; TAP1; TAP2; TNFRSF13B
KEGG: Legionellosis	2.575; 1.944; 2.013	9.579; 5.673; 3.03		ARF1; BCL2L13; BNIP3; C3; CASP1; CASP3; CASP7; CASP8; CASP9; CD14; CLK1; CR1; CXCL1; CXCL8; EEF1G; HBS1L; HSF1; HSPA1A; HSPA1B; HSPA1L; HSPA6; HSPA8; HSPD1; IL18; IL1B; ITGAM; ITGB2; MYD88; NAIP; NFKB1; NFKB2; NFKBIA; NLRC4; PYCARD; RAB1A; RAB1B; SAR1A; SAR1B; SEC22B; TLR2; TLR4; TLR5; TNF; VCP
KEGG: Epithelial cell signalling in Helicobacter pylori infection	2.472; 4.05; 3.356		4.263; 3.426; 5.376	ADAM10; ADAM17; ATP6AP1; ATP6V0A2; ATP6V0B; ATP6V0C; ATP6V0D1; ATP6V0E1; ATP6V0E2; ATP6V1A; ATP6V1B2; ATP6V1C1; ATP6V1D; ATP6V1E1; ATP6V1E2; ATP6V1F; ATP6V1G1; ATP6V1H; CASP3; CCL5; CDC42; CHUK; CSK; CXCL1; CXCL8; CXCR1; CXCR2; F11R; HBEGF; IKBKB; IKBKG; JUN; LYN; MAP2K4; MAP3K14; MAPK13; MAPK14; NFKB1; NFKBIA; PAK1; PLCG2; PTPN11; TCIRG1
KEGG: Endocrine resistance	2.208; 2.692; 4.027			ADCY3; ADCY4; ADCY7; AKT1; AKT2; ARAF; BAD; BAX; BIK; BRAF; CARM1; CDK4; CDKN1A; CDKN1B; CDKN2A; CDKN2C; E2F1; E2F3; FOS; GNAS; GPER1; GRB2; HBEGF; IGF1R; JAG2; JUN; KRAS; MAP2K1; MAP2K2; MAPK13; MAPK14; MAPK3; MED1; MMP9; NCOR1; NOTCH2; NRAS; PIK3CA; PIK3CB; PIK3CD; PIK3R1; PRKACB; RAF1; RB1; RPS6KB1; RPS6KB2; SHC1; SOS1; SOS2; SP1
KEGG: Rheumatoid arthritis	1.96; 2.399; 2.685		8.817; 8.538; 5.376	ACP5; ATP6AP1; ATP6V0A2; ATP6V0B; ATP6V0C; ATP6V0D1; ATP6V0E1; ATP6V0E2; ATP6V1A; ATP6V1B2; ATP6V1C1; ATP6V1D; ATP6V1E1; ATP6V1E2; ATP6V1F; ATP6V1G1; ATP6V1H; CCL2; CCL20; CCL3; CCL3L1; CCL5; CD86; CTSK; CTSL; CXCL1; CXCL5; CXCL8; FOS; HLA-DMA; HLA-DMB; HLA-DOB; HLA-DPA1; HLA-DPB1; HLA-DQA1; HLA-DQB1; HLA-DRA; HLA-DRB1; HLA-DRB5; ICAM1; IFNG; IL15; IL18; IL1A; IL1B; ITGAL; ITGB2; JUN; LTB; MMP1; TCIRG1; TGFB1; TLR2; TLR4; TNF; TNFSF13; TNFSF13B
KEGG: Purine metabolism	1.404; 1.705; 2.685			ADA; ADA2; ADCY3; ADCY4; ADCY7; ADPRM; ADSL; ADSS; AK1; AK2; AK3; AK4; AMPD2; AMPD3; APRT; ATIC; CANT1; DCK; DGUOK; ENPP4; ENTPD1; ENTPD4; ENTPD5; ENTPD6; FHIT; GART; GMPR; GMPR2; GMPS; GUCY1A1; GUCY1B1; GUCY2D; GUK1; HDDC3; HPRT1; IMPDH1; IMPDH2; ITPA; NME1; NME1-NME2; NME3; NME4; NME6; NT5C; NT5C2; NT5C3A; NT5M; NTPCR; NUDT16; NUDT5; NUDT9; PAICS; PAPSS1; PAPSS2; PDE1B; PDE2A; PDE3B; PDE4A; PDE4B; PDE4D; PDE6D; PDE6G; PDE7A; PDE8A; PGM1; PGM2; PKM; PNP; PPAT; PRPS1; PRPS2; PRUNE1; RRM1; RRM2; RRM2B

36

Pathway				Genes
KEGG: Salivary secretion	0.979; 1.44; 2.013	5.848; 3.685; 6.061	0.548; 1.377; 2.151	ADCY3; ADCY4; ADCY7; ADRB2; AMY1A; AQP5; ATP1A1; ATP1B1; ATP1B3; ATP2B1; ATP2B4; BST1; CALM1; CALM2; CALML4; CAMP; CD38; CST2; CST3; CST4; GNAQ; GNAS; GUCY1A1; GUCY1B1; KCNMA1; KCNN4; LYZ; PLCB1; PLCB2; PRH1; PRKACB; PRKCB; SLC9A1
KEGG: Progesterone-mediated oocyte maturation	0.739; 0.379; 0.671			ADCY3; ADCY4; ADCY7; AKT1; AKT2; ANAPC1; ANAPC11; ANAPC13; ANAPC4; ANAPC5; ARAF; AURKA; BRAF; BUB1; CCNA1; CCNA2; CCNB1; CCNB2; CDC16; CDC25B; CDC26; CDC27; CDK1; CDK2; CPEB2; CPEB3; CPEB4; GNAI2; GNAI3; HSP90AA1; HSP90AB1; IGF1R; KIF22; KRAS; MAD1L1; MAD2L1; MAD2L2; MAP2K1; MAPK13; MAPK14; MAPK3; PDE3B; PIK3CA; PIK3CB; PIK3CD; PIK3R1; PKMYT1; PRKACB; RAF1; RPS6KA1; RPS6KA2; RPS6KA3; SPDYE1; STK10
CORUM: 12S U11 snRNP	0.68; 1.354; 0.671		5.814; 4.818; 5.376	PDCD7; PRKRIP1; SNRNP25; SNRNP35; SNRNP48; SNRPB; SNRPD1; SNRPD2; SNRPD3; SNRPE; SNRPF; SNRPG; SRSF7; YBX1
KEGG: Cytosolic DNA-sensing pathway	0.621; 0.525; 1.342		7.675; 5.934; 4.301	ADAR; AIM2; CASP1; CCL4; CCL4L1; CCL5; CGAS; CHUK; CXCL10; DDX58; IKBKB; IKBKE; IKBKG; IL18; IL1B; IRF3; IRF7; MAVS; NFKB1; NFKBIA; POLR1C; POLR1D; POLR2E; POLR2F; POLR2H; POLR2K; POLR2L; POLR3C; POLR3D; POLR3E; POLR3GL; POLR3K; PYCARD; RIPK1; RIPK3; TBK1; TMEM173; TREX1; ZBP1
KEGG: Glutathione metabolism	0.326; 0.323; 0.671			ANPEP; CHAC2; G6PD; GCLC; GCLM; GGT1; GPX1; GPX4; GPX7; GSR; GSS; GSTK1; GSTM3; GSTO1; GSTP1; HPGDS; IDH1; IDH2; LAP3; MGST1; MGST2; MGST3; NAT8B; ODC1; OPLAH; PGD; RRM1; RRM2; RRM2B; SMS; SRM; TXNDC12
KEGG: Steroid biosynthesis		9.044; 9.84; 9.091		CYP2R1; CYP51A1; DHCR24; DHCR7; EBP; FDFT1; LIPA; MSMO1; NSDHL; SOAT1; SQLE; TM7SF2
KEGG: RNA polymerase		6.459; 3.452; 6.061		POLR1B; POLR1C; POLR1D; POLR1E; POLR2A; POLR2B; POLR2C; POLR2E; POLR2F; POLR2G; POLR2H; POLR2I; POLR2J; POLR2J3; POLR2K; POLR2L; POLR3C; POLR3D; POLR3E; POLR3GL; POLR3K; ZNRD1
KEGG: mTOR signalling pathway		5.646; 5.038; 3.03		AKT1; AKT1S1; AKT2; ATP6V1A; ATP6V1B2; ATP6V1C1; ATP6V1D; ATP6V1E1; ATP6V1E2; ATP6V1F; ATP6V1G1; ATP6V1H; BRAF; CAB39; CASTOR1; CHUK; CLIP1; DDIT4; DVL1; EIF4B; EIF4E2; EIF4EBP1; FLCN; FNIP2; FZD1; FZD2; GRB10; GRB2; GSK3B; IGF1R; IKBKB; INSR; KRAS; LAMTOR1; LAMTOR2; LAMTOR3; LAMTOR4; MAP2K1; MAP2K2; MAPK3; MAPKAP1; MIOS; MLST8; NPRL2; NPRL3; NRAS; PDPK1; PIK3CA; PIK3CB; PIK3CD; PIK3R1; PRKAA1; PRKCB; PRR5; PTEN; RAF1; RHEB; RHOA; RICTOR; RPS6; RPS6KA1; RPS6KA2; RPS6KA3; RPS6KB1; RPS6KB2; RRAGA; RRAGB; RRAGC; RRAGD; SEC13; SESN2; SGK1; SLC38A9; SLC3A2; SOS1; SOS2; STK11; STRADA; STRADB; TBC1D7; TELO2; TNF; TNFRSF1A; TSC2; TTI1; ULK1; WDR59; WNT3
CORUM: SMN complex		3.084; 2.097; 3.03		DDX20; GEMIN4; GEMIN7; GEMIN8; SNRPB; SNRPD1; SNRPD2; SNRPD3; SNRPE; SNRPF; SNRPG; STRAP
KEGG: Ribosome biogenesis in eukaryotes		1.874; 3.782; 3.03		BMS1; CSNK2A1; CSNK2A2; CSNK2B; DKC1; DROSHA; EIF6; EMG1; FBL; FCF1; GNL2; GNL3; GNL3L; GTPBP4; HEATR1; IMP3; IMP4; LSG1; MPHOSPH10; NHP2; NOB1; NOP10; NOP56; NOP58; NXF1; NXT1; NXT2; POP4; POP5; POP7; PWP2; RAN; RCL1; REXO1; REXO2; RPP25L; RPP38; RRP7A; SBDS; SNU13; TCOF1; UTP14A; UTP14C; UTP18; UTP4; UTP6; WDR36; WDR75; XPO1; XRN1; XRN2

Pathway				Genes
KEGG: Prion diseases		0.571; 2.383; 3.03		BAX; C1QA; C1QB; C1QC; C5; CCL5; EGR1; ELK1; FYN; HSPA1A; HSPA5; IL1A; IL1B; LAMC1; MAP2K1; MAP2K2; MAPK3; PRKACB; PRNP; SOD1; STIP1
CORUM: Kinase maturation complex 1		0.228; 2.612; 3.03		CDC37; HSP90AA1; HSP90AB1; HSPA4; MAP2K5; MAP3K3; PDRG1; PFDN2; TRAF7; YWHAB; YWHAG; YWHAH; YWHAQ; YWHAZ
KEGG: Arachidonic acid metabolism			8.65; 10.792; 11.828	AKR1C3; ALOX12; ALOX5; CBR1; CYP4A22; CYP4F2; CYP4F3; GGT1; GPX1; GPX7; HPGDS; JMJD7-PLA2G4B; LTA4H; LTC4S; PLA2G12A; PLA2G4A; PLA2G4F; PLB1; PRXL2B; PTGDS; PTGES; PTGES3; PTGS1; PTGS2; TBXAS1
KEGG: Collecting duct acid secretion			5.466; 7.8; 8.602	ATP6V0A2; ATP6V0C; ATP6V0D1; ATP6V0E1; ATP6V0E2; ATP6V1A; ATP6V1B2; ATP6V1C1; ATP6V1D; ATP6V1E1; ATP6V1E2; ATP6V1F; ATP6V1G1; CA2; SLC4A1; TCIRG1
KEGG: Regulation of lipolysis in adipocytes			5.229; 3.744; 5.376	ABHD5; ADCY3; ADCY4; ADCY7; ADRB2; AKT1; AKT2; GNAI2; GNAI3; GNAS; INSR; IRS2; MGLL; NPPA; PDE3B; PIK3CA; PIK3CB; PIK3CD; PIK3R1; PRKACB; PTGS1; PTGS2
KEGG: Leishmaniasis			4.508; 6.629; 6.452	C3; CR1; CYBA; CYBB; ELK1; FCGR1A; FCGR2A; FCGR2C; FCGR3A; FCGR3B; FOS; HLA-DMA; HLA-DMB; HLA-DOB; HLA-DPA1; HLA-DPB1; HLA-DQA1; HLA-DQB1; HLA-DRA; HLA-DRB1; HLA-DRB5; IFNG; IFNGR1; IFNGR2; IL10; IL1A; IL1B; IRAK1; IRAK4; ITGA4; ITGAM; ITGB1; ITGB2; JAK1; JAK2; JUN; MAP3K7; MAPK13; MAPK14; MAPK3; MARCKSL1; MYD88; NCF1; NCF2; NCF4; NFKB1; NFKBIA; PRKCB; PTGS2; PTPN6; STAT1; TAB1; TAB2; TGFB1; TLR2; TLR4; TNF; TRAF6

References

1. Trifonov V, Khiabanian H, Rabadan R: **Geographic dependence, surveillance, and origins of the 2009 influenza A (H1N1) virus.** *N Engl J Med* 2009, **361:**115-119.

2. Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, Haagmans BL, Lauber C, Leontovich AM, Neuman BW, et al: **The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2.** *Nat Microbiol* 2020, **5:**536-544.

3. Alhazzani W, Moller MH, Arabi YM, Loeb M, Gong MN, Fan E, Oczkowski S, Levy MM, Derde L, Dzierba A, et al: **Surviving Sepsis Campaign: guidelines on the management of critically ill adults with Coronavirus Disease 2019 (COVID-19).** *Intensive Care Med* 2020.

4. van Asten L, Luna Pinzon A, de Lange DW, de Jonge E, Dijkstra F, Marbus S, Donker GA, van der Hoek W, de Keizer NF: **Estimating severity of influenza epidemics from severe acute respiratory infections (SARI) in intensive care units.** *Crit Care* 2018, **22:**351.

5. Angus DC, van der Poll T: **Severe sepsis and septic shock.** *N Engl J Med* 2013, **369:**840-851.

6. Liu V, Escobar GJ, Greene JD, Soule J, Whippy A, Angus DC, Iwashyna TJ: **Hospital deaths in patients with sepsis from 2 independent cohorts.** *JAMA* 2014, **312:**90-92.

7. Cohen J, Vincent JL, Adhikari NK, Machado FR, Angus DC, Calandra T, Jaton K, Giulieri S, Delaloye J, Opal S, et al: **Sepsis: a roadmap for future research.** *Lancet Infect Dis* 2015, **15:**581-614.

8. Scicluna BP, van Vught LA, Zwinderman AH, Wiewel MA, Davenport EE, Burnham KL, Nurnberg P, Schultz MJ, Horn J, Cremer OL, et al: **Classification of patients with sepsis according to blood genomic endotype: a prospective cohort study.** *Lancet Respir Med* 2017, **5:**816-826.

9. Hu Y, Cheng L, Zhong W, Chen M, Zhang Q: **Bioinformatics Analysis of Gene Expression Profiles for Risk Prediction in Patients with Septic Shock.** *Med Sci Monit* 2019, **25:**9563-9571.

10. Maslove DM, Wong HR: **Gene expression profiling in sepsis: timing, tissue, and translational considerations.** *Trends Mol Med* 2014, **20:**204-213.

11. Scicluna BP, Klein Klouwenberg PM, van Vught LA, Wiewel MA, Ong DS, Zwinderman AH, Franitza M, Toliat MR, Nurnberg P, Hoogendijk AJ, et al: **A**

molecular biomarker to diagnose community-acquired pneumonia on intensive care unit admission.** *Am J Respir Crit Care Med* 2015, **192:**826-835.

12. McHugh L, Seldon TA, Brandon RA, Kirk JT, Rapisarda A, Sutherland AJ, Presneill JJ, Venter DJ, Lipman J, Thomas MR, et al: **A Molecular Host Response Assay to Discriminate Between Sepsis and Infection-Negative Systemic Inflammation in Critically Ill Patients: Discovery and Validation in Independent Cohorts.** *PLoS Med* 2015, **12:**e1001916.

13. Sweeney TE, Shidham A, Wong HR, Khatri P: **A comprehensive time-course-based multicohort analysis of sepsis and sterile inflammation reveals a robust diagnostic gene set.** *Sci Transl Med* 2015, **7:**287ra271.

14. Parlato M, Cavaillon JM: **Host response biomarkers in the diagnosis of sepsis: a general overview.** *Methods Mol Biol* 2015, **1237:**149-211.

15. Sweeney TE, Wong HR: **Risk Stratification and Prognosis in Sepsis: What Have We Learned from Microarrays?** *Clin Chest Med* 2016, **37:**209-218.

16. Peluffo AE: **The "Genetic Program": Behind the Genesis of an Influential Metaphor.** *Genetics* 2015, **200:**685-696.

17. Raser JM, O'Shea EK: **Noise in gene expression: origins, consequences, and control.** *Science* 2005, **309:**2010-2013.

18. Sato K, Ito Y, Yomo T, Kaneko K: **On the relation between fluctuation and response in biological systems.** *Proc Natl Acad Sci U S A* 2003, **100:**14086-14090.

19. Lehner B, Kaneko K: **Fluctuation and response in biology.** *Cell Mol Life Sci* 2011, **68:**1005-1010.

20. de Jong TV, Moshkin YM, Guryev V: **Gene expression variability: the other dimension in transcriptome analysis.** *Physiol Genomics* 2019, **51:**145-158.

21. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25:**25-29.

22. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M: **Data, information, knowledge and principle: back to metabolism in KEGG.** *Nucleic Acids Res* 2014, **42:**D199-205.

23. Giurgiu M, Reinhard J, Brauner B, Dunger-Kaltenbach I, Fobo G, Frishman G, Montrone C, Ruepp A: **CORUM: the comprehensive resource of mammalian protein complexes-2019.** *Nucleic Acids Res* 2019, **47:**D559-D563.

24.   Pan KH, Lih CJ, Cohen SN: **Effects of threshold choice on biological conclusions reached during analysis of gene expression by DNA microarrays.** *Proc Natl Acad Sci U S A* 2005, **102:**8961-8965.

25.   Simillion C, Liechti R, Lischer HE, Ioannidis V, Bruggmann R: **Avoiding the pitfalls of gene set enrichment analysis with SetRank.** *BMC Bioinformatics* 2017, **18:**151.

26.   McCarthy DJ, Chen Y, Smyth GK: **Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation.** *Nucleic Acids Res* 2012, **40:**4288-4297.

27.   Zhao S, Fung-Leung WP, Bittner A, Ngo K, Liu X: **Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells.** *PLoS One* 2014, **9:**e78644.

28.   Bermejo-Martin JF, Martin-Loeches I, Rello J, Anton A, Almansa R, Xu L, Lopez-Campos G, Pumarola T, Ran L, Ramirez P, et al: **Host adaptive immunity deficiency in severe pandemic influenza.** *Crit Care* 2010, **14:**R167.

29.   Rigby RA, Stasinopoulos DM: **Generalized Additive Models for Location, Scale and Shape.** *Journal of the Royal Statistical Society Series C (Applied Statistics)* 2005, **54:**507-554.

30.   Brealey D, Singer M: **Mitochondrial Dysfunction in Sepsis.** *Curr Infect Dis Rep* 2003, **5:**365-371.

31.   Sharma LK, Lu J, Bai Y: **Mitochondrial respiratory complex I: structure, function and implication in human diseases.** *Curr Med Chem* 2009, **16:**1266-1277.

32.   Tretter L, Horvath G, Holgyesi A, Essek F, Adam-Vizi V: **Enhanced hydrogen peroxide generation accompanies the beneficial bioenergetic effects of methylene blue in isolated brain mitochondria.** *Free Radic Biol Med* 2014, **77:**317-330.

33.   Gureev AP, Shaforostova EA, Popov VN, Starkov AA: **Methylene blue does not bypass Complex III antimycin block in mouse brain mitochondria.** *FEBS Lett* 2019, **593:**499-503.

34.   Kwok ES, Howes D: **Use of methylene blue in sepsis: a systematic review.** *J Intensive Care Med* 2006, **21:**359-363.

35.   Jang DH, Nelson LS, Hoffman RS: **Methylene blue for distributive shock: a potential new use of an old antidote.** *J Med Toxicol* 2013, **9:**242-249.

36.   Chen T, Guestrin C: **XGBoost: A Scalable Tree Boosting System.** In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data*

Mining. pp. 785–794. San Francisco, California, USA: Association for Computing Machinery; 2016:785–794.

37. Valencia-Zapata GA, Mejia D, Klimeck G, Zentner M, Ersoy O: **A Statistical Approach to Increase Classification Accuracy in Supervised Learning Algorithms.** In *arXiv e-prints*. pp. arXiv:1709.01439; 2017:arXiv:1709.01439.

38. Demello D, Kierol-Andrews L, Scalise PJ: **Severe sepsis and acute respiratory distress syndrome from community-acquired legionella pneumonia: case report.** *Am J Crit Care* 2007, **16:**320, 317.

39. Han XY, Tarrand JJ, Dickey BF, Esteva FJ: **Helicobacter pylori bacteremia with sepsis syndrome.** *J Clin Microbiol* 2010, **48:**4661-4663.

40. Endris M, Takele Y, Woldeyohannes D, Tiruneh M, Mohammed R, Moges F, Lynen L, Jacobs J, van Griensven J, Diro E: **Bacterial sepsis in patients with visceral leishmaniasis in Northwest Ethiopia.** *Biomed Res Int* 2014, **2014:**361058.

41. Blake WJ, Balazsi G, Kohanski MA, Isaacs FJ, Murphy KF, Kuang Y, Cantor CR, Walt DR, Collins JJ: **Phenotypic consequences of promoter-mediated transcriptional noise.** *Mol Cell* 2006, **24:**853-865.

42. Fernandes D, Assreuy J: **Nitric oxide and vascular reactivity in sepsis.** *Shock* 2008, **30 Suppl 1:**10-13.

43. Vasko R: **Peroxisomes and Kidney Injury.** *Antioxid Redox Signal* 2016, **25:**217-231.

44. Chen X, Wang Y, Xie X, Chen H, Zhu Q, Ge Z, Wei H, Deng J, Xia Z, Lian Q: **Heme Oxygenase-1 Reduces Sepsis-Induced Endoplasmic Reticulum Stress and Acute Lung Injury.** *Mediators Inflamm* 2018, **2018:**9413876.

45. Wei H, Kemp S, McGuinness MC, Moser AB, Smith KD: **Pharmacological induction of peroxisomes in peroxisome biogenesis disorders.** *Ann Neurol* 2000, **47:**286-296.

46. Carvalho BS, Irizarry RA: **A framework for oligonucleotide microarray preprocessing.** *Bioinformatics* 2010, **26:**2363-2367.

47. Filzmoser P, Maronna R, Werner M: **Outlier identification in high dimensions.** *Computational Statistics & Data Analysis* 2008, **52:**1694-1711.

48. Bolstad BM, Irizarry RA, Astrand M, Speed TP: **A comparison of normalization methods for high density oligonucleotide array data based on variance and bias.** *Bioinformatics* 2003, **19:**185-193.

49. R Core Team: **R: A Language and Environment for Statistical Computing.** Vienna, Austria: R Foundation for Statistical Computing; 2013.

50.   Kloke JD, McKean JW, Rashid MM: **Rank-Based Estimation and Associated Inferences for Linear Models With Cluster Correlated Errors.** *Journal of the American Statistical Association* 2009, **104:**384-390.

51.   Fitzpatrick SF: **Immunometabolism and Sepsis: A Role for HIF?** *Front Mol Biosci* 2019, **6:**85.

52.   Vanderhaeghen T, Vandewalle J, Libert C: **Hypoxia-inducible factors in metabolic reprogramming during sepsis.** *FEBS J* 2020.

53.   Choi ME, Price DR, Ryter SW, Choi AMK: **Necroptosis: a crucial pathogenic mediator of human disease.** *JCI Insight* 2019, **4**.

54.   Sauler M, Bazan IS, Lee PJ: **Cell Death in the Lung: The Apoptosis-Necroptosis Axis.** *Annu Rev Physiol* 2019, **81:**375-402.

55.   Zhou H, Coveney AP, Wu M, Huang J, Blankson S, Zhao H, O'Leary DP, Bai Z, Li Y, Redmond HP, et al: **Activation of Both TLR and NOD Signaling Confers Host Innate Immunity-Mediated Protection Against Microbial Infection.** *Front Immunol* 2018, **9:**3082.

56.   Yang YX, Li L: **Identification of potential biomarkers of sepsis using bioinformatics analysis.** *Exp Ther Med* 2017, **13:**1689-1696.

57.   Pinheiro da Silva F, Aloulou M, Skurnik D, Benhamou M, Andremont A, Velasco IT, Chiamolera M, Verbeek JS, Launay P, Monteiro RC: **CD16 promotes Escherichia coli sepsis through an FcR gamma inhibitory pathway that prevents phagocytosis and facilitates inflammation.** *Nat Med* 2007, **13:**1368-1374.

58.   Yin X, Xin H, Mao S, Wu G, Guo L: **The Role of Autophagy in Sepsis: Protection and Injury to Organs.** *Front Physiol* 2019, **10:**1071.

59.   Su L, Li H, Xie A, Liu D, Rao W, Lan L, Li X, Li F, Xiao K, Wang H, et al: **Dynamic changes in amino acid concentration profiles in patients with sepsis.** *PLoS One* 2015, **10:**e0121933.

60.   Jung WJ, Park BH, Chung KS, Kim SY, Kim EY, Jung JY, Kang YA, Kim YS, Kim SK, Chang J, Park MS: **Glucagon levels, disease severity, and outcome in severe sepsis.** *Shock* 2015, **43:**563-568.

61.   Baumgartner MR, Horster F, Dionisi-Vici C, Haliloglu G, Karall D, Chapman KA, Huemer M, Hochuli M, Assoun M, Ballhausen D, et al: **Proposed guidelines for the diagnosis and management of methylmalonic and propionic acidemia.** *Orphanet J Rare Dis* 2014, **9:**130.

62. Truong KK, Lam MT, Grandner MA, Sassoon CS, Malhotra A: **Timing Matters: Circadian Rhythm in Sepsis, Obstructive Lung Disease, Obstructive Sleep Apnea, and Cancer.** *Ann Am Thorac Soc* 2016, **13:**1144-1154.

63. Feketeova E, Li Z, Joseph B, Shah R, Spolarics Z, Ulloa L: **Dopaminergic Control of Inflammation and Glycemia in Sepsis and Diabetes.** *Front Immunol* 2018, **9:**943.

64. Pacheco R, Contreras F, Zouali M: **The dopaminergic system in autoimmune diseases.** *Front Immunol* 2014, **5:**117.

65. Lee CW, Chen HJ, Liang JA, Kao CH: **Risk of sepsis in patients with amyotrophic lateral sclerosis: a population-based retrospective cohort study in Taiwan.** *BMJ Open* 2017, **7:**e013761.

66. Mukhopadhyay S, Thatoi PK, Pandey AD, Das BK, Ravindran B, Bhattacharjee S, Mohapatra SK: **Transcriptomic meta-analysis reveals up-regulation of gene expression functional in osteoclast differentiation in human septic shock.** *PLoS One* 2017, **12:**e0171689.

67. Yoseph BP, Klingensmith NJ, Liang Z, Breed ER, Burd EM, Mittal R, Dominguez JA, Petrie B, Ford ML, Coopersmith CM: **Mechanisms of Intestinal Barrier Dysfunction in Sepsis.** *Shock* 2016, **46:**52-59.