(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 December 2012 (20.12.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/174129 Al

(51) **International Patent Classification:**
C12Q 1/68 (2006.01)

(21) **International Application Number:**
PCT/US20 12/042275

(22) **International Filing Date:**
13 June 2012 (13.06.2012)

(25) **Filing Language:** English

(26) **Publication Language:** English

(30) **Priority Data:**
61/496,684 14 June 201 1 (14.06.201 1) US

(71) **Applicants** *(for all designated States except US):* **THE BOARD OF TRUSTEES OF THE LELAND STAN¬ FORD JUNIOR UNIVERSITY** [US/US]; 1705 El Camino Real, Palo Alto, California 94306-1 106 (US). **RE¬ SEARCH INSTITUTE AT NATIONWIDE CHIL¬ DREN'S HOSPITAL** [US/US]; 700 Children's Drive, Columbus, Ohio 43205 (US).

(72) **Inventors; and**

(75) **Inventors/Applicants** *(for US only):* **LING, Bruce Xuefeng** [US/US]; 1400 Waverley Street, Palo Alto, California 94301 (US). **SYLVESTER, Karl G.** [US/US]; 1230 Patlen Drive, Los Altos, California 94022 (US). **MOSS, R. Lawrence** [US/US]; 7347 Lambton Park Road, New A1bany, OH 43054-8893 (US).

(74) **Agent: ALCAMO, Elizabeth A.;** Bozicevic, Field & Francis LLP, 1900 University Avenue, Suite 200, East Palo Alto, California 94303 (US).

(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available):* AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available):* ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

WO 2012/174129 A1

(54) Title: BIOMARKERS FOR NECROTIZING ENTEROCOLITIS AND SEPSIS

(57) Abstract: Aspects of the invention include methods, compositions, and kits for diagnosing Necrotizing Enterocolitis (NEC), for diagnosing sepsis, for providing a prognosis for a patient with NEC, and for predicting responsiveness of a patient with NEC to medical intervention. These methods find use in a number of applications, such as diagnosing and treating infants who are suspected of having NEC, intestinal perforation (IP), or sepsis.

BIOMARKERS FOR NECROTIZING ENTEROCOLITIS AND SEPSIS

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] Pursuant to 35 U.S.C. § 119 (e), this application claims priority to the filing date of the United States Provisional Patent Application Serial No. 61/496,684 filed June 14, 2011; the disclosure of which is herein incorporated by reference.

FIELD OF THE INVENTION

[0002] This invention pertains to the fields of necrotizing enterocolitis and sepsis.

BACKGROUND OF THE INVENTION

[0003] Necrotizing enterocolitis (NEC), intestinal perforation (IP) and sepsis are three life-threatening gastrointestinal diseases among neonates and together constitute a leading cause of overall morbidity and mortality in premature newborns. However, there is considerable overlap in the early clinical presentation of NEC, IP and sepsis in newborns. Furthermore, while half of NEC-affected infants will recover with medical therapy alone (the M class), 30-50% develop a progressive form of the disease (Progressive Necrotizing Enterocolitis) that requires surgery (the S class) to prevent mortality. Currently utilized clinical parameters including laboratory tests and diagnostic imaging fail to capture the nuanced differences between these entities during their onset and progression. Protein biomarkers detectable in clinically available specimens would provide the needed molecular diagnostic and prognostic "fingerprint" against which we can begin to measure various interventions. Such biomarkers could be used to improved methods for diagnostic and prognostic class prediction in NEC, IP and sepsis, and to improve predictions on responsiveness to known and new therapies. The present invention addresses these issues.

Publications

[0004] U.S. Application No. 2009/0191551 teaches using the level of secretor antigens in a biological fluid as a marker to predict the risk of developing NEC. Thuijls G, et al. (2010) *Non-invasive markers for early diagnosis and determination of the severity of necrotizing enterocolitis.* Ann Surg. 251(6):1174-80, discusses using l-FABP and claudin-3 protein levels in urine and calprotectin protein levels in fecal matter as diagnostic markers of NEC, and I-FABP protein levels in urine as a prognostic marker of disease severity. Evennett N, et al. (2009) *A systematic review of serologic tests in the diagnosis of necrotizing enterocolitis.* J Pediatr Surg. 44(11):2192-201 is a review of publications that were deemed by the authors to be potentially relevant to diagnostic performance of serological tests in NEC. Young C, et al. (2009) *Biomarkers for infants at risk for necrotizing enterocolitis: clues to prevention?* Pediatr Res. 65(5 Pt 2):91R-97R is a review that discusses the potential value of genomic and

proteomic studies of NEC in the identification of biomarkers for early diagnosis and targeted prevention of this disease.

SUMMARY OF THE INVENTION

[0005] Aspects of the invention include methods, compositions, and kits for diagnosing Necrotizing Enterocolitis (NEC), for diagnosing sepsis, for providing a prognosis for a patient with NEC, and for predicting responsiveness of a patient with NEC to medical intervention. These methods find use in a number of applications, such as diagnosing and treating infants who are suspected of having NEC, intestinal perforation (IP), or sepsis. These and other objects, advantages, and features of the invention will become apparent to those persons skilled in the art upon reading the details of the compositions and methods as more fully described below.

[0006] In some aspects of the invention, methods are provided for diagnosing NEC. In these methods, a NEC-Dx expression signature is obtained for a patient, where a NEC-Dx expression signature comprises the quantitative data on the expression level of one or more NEC-Dx genes, i.e. genes that are expressed at elevated levels in patients with NEC versus unaffected individuals. The NEC-Dx expression signature is then compared to an NEC-Dx expression signature from a reference sample, and the results of this comparison are employed to provide a diagnosis of NEC to the patient. In some embodiments, the patient is suspected of having NEC, intestinal perforation (IP), or sepsis.

[0007] In some embodiments, the NEC-Dx expression signature is obtained by detecting the amount of protein in a body fluid that is encoded by one or more NEC-Dx genes to arrive at a NEC-Dx protein signature. In some embodiments, the body fluid is urine. In some embodiments, the one or more NEC-Dx genes is selected from the group consisting of CD14, SAP1 , PEDF, Q6ZUQ4, OBFC2B, COL1 1A2, NBEAL2, GRASP, HUWE1 , COL1A2, HOXD3, DSG4, KRTAP5-1 1, Y1020, FGA, CTAPIII/PPBP, SAA1 , B2M, TTR, OSTP/OPN, APOA4, C08G, ANGT, FIBA, PROF1 , UMOD, PLSL, and LMAN2, where elevated levels of one or more of these genes is diagnostic for NEC.

[0008] In some embodiments, the method further comprises obtaining an NEC clinical score. In such embodiments, the NEC-Dx expression signature and NEC clinical score are compared to an NEC-Dx expression signature and NEC clinical score from a reference sample, and the results of both comparisons are employed to provide a diagnosis of NEC.

[0009] In some aspects of the invention, methods are provided for diagnosing sepsis in a patient. In these methods, a sepsis-Dx expression signature is obtained from the patient, where a sepsis-Dx expression signature comprises the quantitative data on the expression level of one or more sepsis-Dx genes, i.e. genes that are expressed at elevated levels in patients with sepsis. The sepsis-Dx expression signature is then compared to a sepsis-Dx

expression signature from a reference sample, and the results of this comparison are employed to provide a diagnosis of sepsis to the patient. In some embodiments, the patient is suspected of having NEC, intestinal perforation (IP), or sepsis.

[0010] In some embodiments, the sepsis-Dx expression signature is obtained by detecting/measuring the amount of protein in a body fluid that is encoded by sepsis-Dx genes to arrive at a sepsis-Dx protein signature. In some embodiments, the body fluid is urine. In some embodiments, the one or more sepsis-Dx genes are selected from the group consisting of ftsy, PROC, MAPI B and CSN5, where elevated levels of one or more of these genes is diagnostic of sepsis.

[0011] In some embodiments, a sepsis clinical score is also obtained, the sepsis-Dx signature and the sepsis clinical score are compared to a sepsis-Dx signature and a sepsis clinical score from a reference sample, and the results of both comparisons are employed to provide a sepsis diagnosis to the patient.

[0012] In some aspects of the invention, methods are provided for providing a prognosis for a patient with NEC, or for predicting responsiveness of an NEC patient to medical therapy versus surgical intervention. In these methods, an NEC-M/S expression signature is obtained for the patient, where the NEC-M/S expression signature comprises quantitative data on the level in a body fluid of proteins encoded by one or more NEC-S genes and/or one or more NEC-M genes, NEC-S genes being genes that are expressed at a higher level in S-class NEC patients (patients that require surgery for recovery) than M-class patients (patients that respond to medical therapy), and NEC-M genes being genes that are expressed at a higher level in M-class NEC patients than S-class patients. The NEC-M/S expression signature is then compared to an NEC-M/S expression signature from a reference sample, and the results of this comparison are employed to provide a prognosis for the patient or to predict the responsiveness of the patient to medical therapy. In some embodiments, the method also provides for making a diagnosis of NEC. In other embodiments, the patient is known to have NEC prior to performing the method.

[0013] In some embodiments, the body fluid is urine. In some embodiments, the one or more NEC-M/S genes are selected from the group consisting of Q6ZUQ4, OBFC2B, COL1 1A2, NBEAL2, GRASP, HUWE1 , COL1A2, HOXD3, DSG4, KRTAP5-1 1, Y 1020, FGA, OSTP/OPN, APOA4, C08G, SAP1 , ANGT, CD14, FIBA, PROF1 , PEDF, UMOD, PLSL, and LMAN2. In some embodiments, the NEC-S gene is selected from the group consisting of Q6ZUQ4, OBFC2B, COL1 1A2, NBEAL2, GRASP, HUWE1 , COL1A2, HOXD3, DSG4, KRTAP5-1 1, Y 1020, FGA, OSTP/OPN, APOA4, C08G, SAP1 , ANGT, CD14, FIBA, PROF1 , and PEDF, where elevated levels of one or more of these genes is diagnostic of NEC-S. In some embodiments, the NEC-M gene is selected from the group consisting of UMOD, PLSL, and LMAN2, where elevated levels of one or more of these genes is diagnostic of NEC-M.

[0014] In some embodiments, an NEC clinical score is also obtained. In some such embodiments, the NEC-M/S signature and the NEC clinical score are compared to a NEC-M/S signature and an NEC clinical score from a reference sample, and the results of both comparisons are employed to provide a prognosis to the patient or to predict the responsiveness of the patient to medical treatment.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] The invention is best understood from the following detailed description when read in conjunction with the accompanying drawings. The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee. It is emphasized that, according to common practice, the various features of the drawings are not to-scale. On the contrary, the dimensions of the various features are arbitrarily expanded or reduced for clarity. Included in the drawings are the following figures.

[0016] **Figure 1**. Bell's NEC staging criteria cannot be utilized to predict NEC progression and outcome. Evaluated with the Bell staging criteria, linear discriminant analysis was performed with training data from NEC M (n=30) and S (n=17) samples. The trained LDA model was then tested with testing data from NEC M (n=1 3) and S (n=9) samples. Estimated probabilities for the training (left) and testing data (right) are plotted (panel A). Samples are partitioned by the true class (upper) and predicted class (lower). The classification results from training (panel B) and testing sets (panel C) are shown as 2X2 contingency tables. Fisher exact test was used to measure P values of the 2X2 table. (D) Unsupervised hierarchical clustering trees based on the Bell staging criteria.

[0017] **Figure 2.** Eleven clinical parameters were selected to classify NEC M and S patients by linear discriminant analysis (LDA). (A) 11 clinical parameters (Mann Whitney U test P value < 0.1) and the corresponding absolute value (ABS) of the first linear discriminant (LD1) from the LDA. (B) Using these 11 clinical parameters, a LDA model was trained with NEC M (n=30) and S (n=17) training samples and tested with data from NEC M (n=1 3) and S (n=9) samples. Estimated probabilities for the training (left) and testing data (right) are plotted. Samples are partitioned by the true class (upper) and predicted class (lower). (C, D) The classification results from the training and testing sets are shown as 2X2 contingency tables. Fisher exact test was used to measure P values of the 2X2 tables. (E) ROC analysis of the classification performance of the LDA model of the 11 clinical parameters.

[0018] **Figure 3.** Unsupervised clustering and pathway analyses of the MSMS identified urine peptides differentiating NEC M (n=17) and S (n=1 1) subjects. (A) Heatmap display of unsupervised clustering analyses of expression of the top 473 urine peptides ranked by significant analyses comparing NEC M and S samples. Manual review of the feature clusters

into I, II, III groups. (B). Data mining software (Ingenuity Systems, www.ingenuity.com, CA) was used with these differential urine peptides' parent proteins to identify and calculate the significance of the gene ontology groups and relevant canonical signaling pathways associated with NEC progression. (C) Overlapping urine peptides found differentiating NEC M, S and Post S groups. m/z: Mass to charge ratio. z: Peptide charge. Relative abundance: the nearest shrunken centroid values have been utilized to represented the relative abundance of the peptide biomarkers in either NEC M or S or Post S patient class with the Color Scale conditional formatting. P*: hydroxyproline.

[0019] **Figure 4.** A 36 urine peptide biomarker panel effectively differentiates NEC M from S subjects. (A). Goodness of separation analysis to select a 36-peptide panel (red asterisk labeled) as the optimal urine peptide biomarker panel for the NEC progression analysis. Using 473 MSMS identified urine peptide data from NEC M and S data sets, as indicated, various classifiers of different panel size (feature #) were tested for their goodness of separation between NEC M (green) and quiescence (red) as shown by the box-whisker graphs. Boxes contain the 50% of values falling between the 25th and 75th percentiles; the horizontal line within the box represents the median value and the "whisker" lines extend to the highest and lowest values. (B) Heatmap display of unsupervised clustering analyses of expression of the 36 urine peptide biomarkers. (C) Relative abundance of the 36 urine peptide abundance by the nearest shrunken centroid values in either NEC M or S patient class with the Color Scale conditional formatting. m/z: Mass to charge ratio. z: Peptide charge. P*: hydroxyproline. The significance of each urine peptide biomarker in differentiating NEC M from S groups was quantified by Mann-Whitney U test and Student T test P values. (D) Pathway analysis using Panther database revealed the pathways in which the protein precursors of the 36 urine peptides are involved.

[0020] **Figure 5.** Significant analysis of NEC M and S subjects found a 30-plasma-protein biomarker panel. (A). Goodness of separation analysis to select a panel of 48 spectral peaks (red asterisk labeled) for the NEC progression analysis. Using 1528 different spectra peak data from NEC M and S sets, as indicated, various classifiers of different panel size (feature #) were tested for their goodness of separation between NEC M (green) and NEC S (red) as shown by the box-whisker graphs. Boxes contain the 50% of values falling between the 25th and 75th percentiles; the horizontal line within the box represents the median value and the "whisker" lines extend to the highest and lowest values. (B) Spectral analysis of the 48 spectral peak found 30 unique plasma proteins. Relative abundance of the 30 proteins were represented by the nearest shrunken centroid values in either NEC M or S patient class with the Color Scale conditional formatting. MW: molecular weight. The significance of each plasma protein in differentiating NEC M from S groups was quantified by Mann-Whitney U test and Student T

test P values. (C) Heatmap display of unsupervised clustering analyses of expression of the 30 plasma protein biomarkers.

[0021] **Figure 6.** Performance evaluation in differentiating NEC 13 M and 11 S subject via (A) 11 clinical parameter based biomarker panel; (B) 36 urine peptide based biomarker panel; (C) 30 plasma protein based biomarker panel; (D) an integrative panel combining all 11 clinical parameters, 36 urine peptides and 30 plasma proteins. Each of the unsupervised clustering results of the NEC M and S subjects are shown as a 2X2 contingency table. Fisher exact test was used to measure P value quantifying the biomarker panel's capability in NEC progression prediction.

[0022] **Figure 7.** Analysis integrating clinical, urine peptide and plasma protein panels derived a biomarker panel of 15 urine peptides and 3 plasma proteins, that predicts NEC progression with high sensitivity and specificity. (A). Goodness of separation and (B) false discovery rate (FDR) analyses chose 18 features from a total of 77 biomarkers (11 clinical parameters, 36 urine peptides and 30 plasma proteins) as the optimal biomarker panel for NEC progression prediction. (C) Relative abundance of the 15 urine peptide and 3 plasma protein abundance by the nearest shrunken centroid values in either NEC M or S patient class with the Color Scale conditional formatting. For urine peptides, MW=MH+-1=m/z-1 . (D) Heatmap display of unsupervised clustering analyses of expression of the 18 (15 urine peptides and 3 plasma proteins) biomarkers. The clustering result is shown as a 2X2 contingency table. Fisher exact test was used to measure the statistical significance (P value) of the 2X2 table. (E). Supervised LDA analysis classifying NEC M and S subjects. Samples are partitioned by the true class (upper) and predicted class (lower). The LDA classification result is shown as a 2X2 contingency table. Fisher exact test was used to measure the statistical significance (P value) of the 2X2 table. (F) ROC analysis of the integrative biomarker panel in discriminating NEC M and S. AUC: area under the curve. The dotted curve is the vertical average of the 500 bootstrapping ROC curves and the boxes and whiskers plot the vertical spread around the average.

[0023] **Figure 8.** A sequential analysis of the clinical and molecular biomarker classifiers for the prediction of NEC progression. (A) NEC clinical scoring system. The samples (violet red-NEC S, sea green-NEC M), sorted by their clinical NEC scores, were grouped into low, intermediate, and high-risk groups. Each particular sample's risk of being NEC S was quantified as the proportion of all NEC S samples with score less than that sample's clinical score in all NEC S samples. (B) Sequential stratification of the NEC subjects using clinical and molecular based classifiers. The molecular based classification result is shown as a 2X2 contingency table. Fisher exact test was used to measure the statistical significance (P value) of the 2X2 table.

[0024] **Figure 9.** Bottom-up urine proteomics discovered am eleven-protein biomarker panel effectively discriminate NEC M from S subjects. We have collected 71 NEC (47 M and 24 S) urine samples to subject them to mass spectrometry (MS) based urine proteome profiling using a bottom up approach. Each proteome was fragmented by trypsin digestion. Full mass spectrometry scan was acquired on an LTQ FTMS, which was followed by MS/MS analysis. Protein identification was performed by searching Swiss-Prot database. Quantification of proteins in different samples was done by means of spectral counting, implementing the recent S1N algorithm (Sardiu, 2010). From the MSMS protein identifications, a separate list of proteins was created for each sample, and the lists were then compared to find differential expressed proteins. For any given protein, the significance of the relative abundance between NEC M and S groups was computed by Student's T test. Urine proteins with low P values discriminating NEC and Sepsis were explored by exploratory box-whisker plot analysis.

Figure 10. Statistical analysis of the eleven-urine-protein NEC M/S biomarker panel. (A) The discriminant probabilities for each sample were calculated from the linear discriminant analysis. The maximum estimated probability for each of the wrongly classified samples is marked with an arrow. (B). A modified 2X2 contingency table was used to the calculated the percentage of classification that agreed with clinical diagnosis for the panel. P value was calculated with Fisher's exact test. (C). The discriminant analysis-derived prediction scores for each sample were used to construct a receiver operating characteristic (ROC) curve. 500 testing data sets, generated by bootstrapping, from the NEC and sepsis data were used to derive estimates of standard errors and confidence intervals for our ROC analysis. The plotted ROC curve is the vertical average of the 500 bootstrapping runs, and the box and whisker plots show the vertical spread around the average.

Figure 11. Bottom-up urine proteomics discovered a seven-protein biomarker panel effectively discriminate NEC from Sepsis subjects. We have collected 71 NEC and 13 Sepsis urine samples to subject them to mass spectrometry (MS) based urine proteome profiling using a bottom up approach. Each proteome was fragmented by trypsin digestion. Full mass spectrometry scan was acquired on an LTQ FTMS, which was followed by MS/MS analysis. Protein identification was performed by searching Swiss-Prot database. Quantification of proteins in different samples was done by means of spectral counting, implementing the recent S1N algorithm {Sardiu, 2010}. From the MSMS protein identifications, a separate list of proteins was created for each sample, and the lists were then compared to find differential expressed proteins. For any given protein, the significance of the relative abundance between NEC and Sepsis groups was computed by Student's T test. Urine proteins with low P values discriminating NEC and Sepsis were explored by exploratory box-whisker plot analysis.

[0025] **Figure 12.** Statistical analysis of the seven-urine-protein NEC/sepsis biomarker panel. (A) The discriminant probabilities for each sample were calculated from the linear discriminant

analysis. The maximum estimated probability for each of the wrongly classified samples is marked with an arrow. (B). A modified 2X2 contingency table was used to the calculated the percentage of classification that agreed with clinical diagnosis for the panel. P value was calculated with Fisher's exact test. (C). The discriminant analysis-derived prediction scores for each sample were used to construct a receiver operating characteristic (ROC) curve. 500 testing data sets, generated by bootstrapping, from the NEC and sepsis data were used to derive estimates of standard errors and confidence intervals for our ROC analysis. The plotted ROC curve is the vertical average of the 500 bootstrapping runs, and the box and whisker plots show the vertical spread around the average.

[0026] General methods in molecular and cellular biochemistry can be found in such standard textbooks as Molecular Cloning: A Laboratory Manual, 3rd Ed. (Sambrook et al., HaRBor Laboratory Press 2001); Short Protocols in Molecular Biology, 4th Ed. (Ausubel et al. eds., John Wiley & Sons 1999); Protein Methods (Bollag et al., John Wiley & Sons 1996); Nonviral Vectors for Gene Therapy (Wagner et al. eds., Academic Press 1999); Viral Vectors (Kaplift & Loewy eds., Academic Press 1995); Immunology Methods Manual (I. Lefkovits ed., Academic Press 1997); and Cell and Tissue Culture: Laboratory Procedures in Biotechnology (Doyle & Griffiths, John Wiley & Sons 1998), the disclosures of which are incorporated herein by reference. Reagents, cloning vectors, and kits for genetic manipulation referred to in this disclosure are available from commercial vendors such as BioRad, Stratagene, Invitrogen, Sigma-Aldrich, and ClonTech. Methodologies for the discovery of urinary peptide biomarkers are detailed in X.B. Ling et al., Advances in Clinical Chemistry 51, 181, 2010.

DETAILED DESCRIPTION OF THE INVENTION

[0027] Before the present methods and compositions are described, it is to be understood that this invention is not limited to particular method or composition described, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting, since the scope of the present invention will be limited only by the appended claims.

[0028] Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limits of that range is also specifically disclosed. Each smaller range between any stated value or intervening value in a stated range and any other stated or intervening value in that stated range is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included or excluded in the range, and each range where either, neither or both limits are included in the smaller ranges is also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the

stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the invention.

[0029] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, some potential and preferred methods and materials are now described. All publications mentioned herein are incorporated herein by reference to disclose and describe the methods and/or materials in connection with which the publications are cited. It is understood that the present disclosure supercedes any disclosure of an incorporated publication to the extent there is a contradiction.

[0030] It must be noted that as used herein and in the appended claims, the singular forms "a", "an", and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a cell" includes a plurality of such cells and reference to "the peptide" includes reference to one or more peptides and equivalents thereof, e.g. polypeptides, known to those skilled in the art, and so forth.

[0031] The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention. Further, the dates of publication provided may be different from the actual publication dates which may need to be independently confirmed.

[0032] Aspects of the invention include methods, compositions, and kits for diagnosing Necrotizing Enterocolitis (NEC), for diagnosing sepsis, for providing a prognosis for a patient with NEC, and for predicting responsiveness of a patient with NEC to medical intervention. These methods find use in a number of applications, such as diagnosing and treating infants who are suspected of having NEC, IP, or sepsis. These and other objects, advantages, and features of the invention will become apparent to those persons skilled in the art upon reading the details of the compositions and methods as more fully described below.

[0033] The term Necrotizing Enterocolitis, or NEC, is used herein to describe the gastrointestinal condition in which a segment of the intestine becomes necrotic; in some instances, the intestinal region perforates, causing peritonitis and often free intra-abdominal air. Infection and inflammation of the gut are hallmarks of the condition, along with abdominal distention, blood in the stool, diarrhea, feeding intolerance, lethargy, temperature instability, and vomiting. There are two classes of NEC: M, for "medical", class; and S, for "surgical", class.

[0034] The terms "medical class NEC", "M class NEC", or "non-progressive NEC" are used interchangeably herein to describe the class of NEC that is typically responsive to medical

therapies, e.g. stage I, stage II, and in some instances stage III of Bell's criteria (Table 1 below). Medical therapy includes, for example, broad spectrum antibiotics for 3-14 days, accompanied intravenous fluids, total parenteral fluids (TPN) and NPO (nothing by mouth).

[0035] The terms "surgical class NEC", "S class NEC" or "progressive NEC" are used interchangeably herein to describe the class of NEC that requires surgical intervention, e.g. stage NIB of Bell's criteria (Table 1 below). In this surgery, gangrenous bowel is resected, and ostomies for intestinal stream diversion are created. With resolution of sepsis and peritonitis, intestinal continuity can be reestablished several weeks or months later.

[0036] The terms "focal intestinal perforation" (FIP), "spontaneous intestinal perforation" (SIP), or "intestinal perforation" (IP) are used interchangeably herein to describe an isolated intestinal perforation that, unlike NEC, is not accompanied by gross necrosis of the tissue. In FIP, the gestational age is significantly lower than in NEC (approx. 24 weeks versus 27 weeks for NEC), the incidence of coexistent respiratory distress syndrome (RDS) is higher (88% versus 37% for NEC), and the age of onset is younger (approx. 7.3 days versus approx. 7.9 days for NEC). See, e.g. Okuyama et al. (2002) Pediatr Surg Int 18:704-706, the disclosure of which is incorporated herein by reference.

[0037] The term "sepsis" is used herein to describe a bacterial infection in the context of fever of greater than 38°C (100.4°F). Blood pressure drops, resulting in shock. Major organs and systems, including the kidneys, liver, lungs, and central nervous system, stop functioning normally. Infection is typically confirmed by a blood culture that reveals bacteria, blood gases that reveal acidosis, kidney function tests that are abnormal, a platelet count that is lower than normal, and/or a white blood cell count that is lower or higher than normal. Other indications of sepsis include a blood differential that shows immature white blood cells, the presence of higher than normal amounts of fibrin degradation products in the blood, and a peripheral smear that shows a low platelet count and destruction of red blood cells. The treatment is typically antibiotics delivered intravenously. In infants, sepsis may be classified as "early onset" (within the first 7 days of birth), which usually results from organisms acquired intrapartum, and "late onset" (more than 7 days after birth), in which the infection is usually by organisms from the environment.

[0038] "Diagnosis" as used herein generally includes a prediction of a subject's susceptibility to a disease or disorder, determination as to whether a subject is presently affected by a disease or disorder, and prognosis of a subject affected by a disease or disorder (e.g., identification of disease states, stages of the disease, likelihood that a patient will die from the disease), and the use of therametrics (e.g., monitoring a subject's condition to provide information as to the effect or efficacy of therapy). "Prediction of a subject's responsiveness to treatment" for the disease or disorder generally includes the prediction of responsiveness (e.g.,

positive response, a negative response, no response at all to, e.g., medical treatment, surgical treatment), and prognosis in view of that predicted responsiveness.

[0039] The term "gene product" or "expression product" are used herein to refer to the RNA transcription products (transcripts) of the gene, including mRNA, and the polypeptide translation products of such RNA transcripts. A gene product can be, for example, an unspliced RNA, an mRNA, a splice variant mRNA, a microRNA, a fragmented RNA, a polypeptide, a post-translationally modified polypeptide, a splice variant polypeptide, etc.

[0040] The term "RNA transcript" as used herein refers to the RNA transcription products of a gene, including, for example, mRNA, an unspliced RNA, a splice variant mRNA, a microRNA, and a fragmented RNA.

[0041] The term "polypeptide" as used herein and as it is applied to a gene refers to the amino acid product encoded by a gene, including, for example, full length gene product, splice variants of the full length gene product, and fragments of the gene product, e.g. peptides.

[0042] The term "expression level" as used herein and as it is applied to a gene refers to the amount of a gene product in a sample, e.g. the normalized value determined for the amount of RNA transcribed from a gene, or the normalized value determined for the amount of polypeptide/protein encoded by the gene. Normalization of the expression level(s) of a gene may be by any well-understood method in the art, e.g. by comparison to the expression of a selected housekeeping gene(s), by comparison to the expression of genes across a whole dataset, etc.

[0043] The term "expression signature" is a representation of the expression levels of one or more genes of interest, more usually two or more genes of interest, and comprises the quantitative data on the expression levels of these one or more genes of interest. Examples of expression signatures include expression profiles, e.g. RNA profiles and protein profiles, and expression scores, e.g. RNA scores and protein scores.

[0044] The term "expression profile" as used herein refers to the normalized expression level of one or more genes of interest, more usually two or more genes of interest, in a patient sample. By "RNA expression profile", or simply "RNA profile", of a patient sample it is meant the normalized expression level of the one or more genes in a patient sample as determined by measuring the amount of RNA transcribed from the one or more genes. By "protein expression profile", or simply "protein profile", of a patient sample it is meant the normalized expression level of the one or more genes in a patient sample as determined by measuring the amount of amino acid product encoded by a gene.

[0045] The term "expression score" as used herein refers to a single metric value that represents the sum of the weighted expression levels of one or more genes of interest, more usually two or more genes of interest, in a patient sample. Weighted expression levels are calculated by multiplying the normalized expression level of each gene by its "weight", the

weight of each gene being determined by analysis of a reference dataset, or "training set", e.g. the datasets provided in the examples section below, e.g. by Principle Component Analysis (PCA), Linear discriminant analysis (LDA), Fisher's linear discriminant analysis, and the like, as are known in the art. Thus, for example, when PCA is used, the expression score is the weighted sum of expression levels of the genes of interest in a sample, where the weights are defined by their first principal component as defined by a reference dataset. By "RNA expression score", or simply "RNA score", of a patient sample it is meant the normalized expression level of the one or more genes in a patient sample as determined by measuring the amount of RNA transcribed from the one or more genes. By "protein expression profile", or simply "protein profile", of a patient sample it is meant the normalized expression level of the one or more genes in a patient sample as determined by measuring the amount of amino acid product encoded by a gene.

[0046] An "NEC-Dx gene" is a gene that is differentially expressed in individuals having NEC relative to individuals that are not affected with NEC.

[0047] An "NEC-Dx expression signature", or more simply, "NEC-Dx signature", is a representation of the expression levels of one or more NEC-Dx genes, and comprises the quantitative data on the expression levels of these one or more NEC-Dx genes. An "NEC-Dx RNA signature" comprises the quantitative data on the amount of RNA transcribed by one or more NEC-Dx genes. An "NEC-Dx protein signature" comprises the quantitative data on the amount of polypeptide encoded by the one or more NEC-Dx genes and/or peptides thereof. An NEC-Dx signature may be in the form of an expression profile or an expression score, as discussed above.

[0048] A "sepsis-Dx gene" or "Sepsis Diagnosis gene" is a gene that is differentially expressed in individuals having sepsis relative to individuals that are not affected with sepsis.

[0049] A "sepsis-Dx expression signature", or more simply, a "sepsis-Dx signature", is a representation of the expression levels of one or more sepsis-Dx genes, and comprises the quantitative data on the expression levels of these one or more genes. A "sepsis-Dx RNA signature" comprises the quantitative data on the amount of RNA transcribed by one or more sepsis-Dx genes. A "sepsis-Dx protein signature" comprises the quantitative data on the amount of polypeptide encoded by one or more sepsis genes and/or peptides thereof. A sepsis-Dx signature may be in the form of an expression profile or an expression score, as discussed above.

[0050] An "NEC-M gene" is a gene that is expressed at a higher level, i.e. is upregulated, in M-class NEC patients than in S-class NEC patients.

[0051] An "NEC-M expression signature", or more simply, a "NEC-M signature", is a representation of the expression levels of one or more NEC-M genes, and comprises the quantitative data on the expression levels of these one or more genes. A "NEC-M RNA

signature" comprises the quantitative data on the amount of RNA transcribed by one or more NEC-M genes. A "NEC-M protein signature" comprises the quantitative data on the amount of polypeptide encoded by one or more sepsis genes and/or peptides thereof. A NEC-M signature may be in the form of an expression profile or an expression score, as discussed above.

[0052] An "NEC-S gene" is a gene that is expressed at a higher level, i.e. is upregulated, in S-class NEC patients than in M-class patients.

[0053] An "NEC-S expression signature", or more simply, a "NEC-S signature", is a representation of the expression levels of one or more NEC-S genes, and comprises the quantitative data on the expression levels of these one or more genes. A "NEC-S RNA signature" comprises the quantitative data on the amount of RNA transcribed by one or more NEC-S genes. A "NEC-S protein signature" comprises the quantitative data on the amount of polypeptide encoded by one or more sepsis genes and/or peptides thereof. A NEC-S signature may be in the form of an expression profile or an expression score, as discussed above.

[0054] An "NEC-M/S gene" is a gene that is differentially expressed in individuals having M class NEC relative to S class NEC or vice versa. An NEC-M/S gene may be used to distinguish between M class NEC and S class NEC or vice versa.

[0055] An "NEC-M/S expression signature", or more simply, "NEC-M/S signature", is a representation of the expression levels of one or more NEC M genes and/or NEC S genes, and comprises the quantitative data on the expression levels of these one or more genes. An "NEC-M/S RNA signature" comprises the quantitative data on the amount of RNA transcribed by one or more NEC-M and/or NEC-S genes. An "NEC-M/S protein signature" comprises the quantitative data on the amount of polypeptide encoded by one or more NEC-M and/or NEC-S genes and/or peptides thereof. In some embodiments, the NEC-M/S signature comprises the quantitative data on the expression level of one ore more NEC-M genes. In some embodiments, the NEC-M/S signature comprises the quantitative data on the expression level of one ore more NEC-S genes. In some embodiments, the NEC-M/S signature comprises the quantitative data on the expression level of one ore more NEC M genes and one or more NEC-S genes. An NEC-M/S signature may be in the form of an expression profile or an expression score, as discussed above.

[0056] The term "risk classification" means a level of risk (or likelihood) that a subject will experience a particular clinical outcome. A subject may be classified into a risk group or classified at a level of risk based on the methods of the present disclosure, e.g. high, medium, or low risk. A "risk group" is a group of subjects or individuals with a similar level of risk for a particular clinical outcome. Examples of NEC risk groups include M-class and the S-class.

[0057] The term "hazard ratio" means the effect of an explanatory variable on the hazard, or risk, of an event occuring. For example, using a Cox proportional hazards regression model, if a variable, e.g. an LSC score, is prognostic, its hazard rate is different in patients with a particular prognosis relative to the hazard rate of other subclasses, and the hazard ratio of the gene is not equal to 1.

[0058] The terms "individual," "subject," "host," and "patient," are used interchangeably herein and refer to any mammalian subject for whom diagnosis, treatment, or therapy is desired, particularly humans.

[0059] The terms "treatment", "treating" and the like are used herein to generally mean obtaining a desired pharmacologic and/or physiologic effect. The effect may be prophylactic in terms of completely or partially preventing a disease or symptom thereof and/or may be therapeutic in terms of a partial or complete cure for a disease and/or adverse effect attributable to the disease. "Treatment" as used herein covers any treatment of a disease in a mammal, and includes: (a) preventing the disease from occurring in a subject which may be predisposed to the disease but has not yet been diagnosed as having it; (b) inhibiting the disease, i.e., arresting its development; or (c) relieving the disease, i.e., causing regression of the disease. The therapeutic agent may be administered before, during or after the onset of disease or injury. The treatment of ongoing disease, where the treatment stabilizes or reduces the undesirable clinical symptoms of the patient, is of particular interest. Such treatment is desirably performed prior to complete loss of function in the affected tissues. The subject therapy will desirably be administered during the symptomatic stage of the disease, and in some cases after the symptomatic stage of the disease.

[0060] Methods, compositions and kits are provided for diagnosing Necrotizing Enterocolitis (NEC) and sepsis, for providing a prognosis for a patient with NEC, and for predicting responsiveness of a patient with NEC to medical therapy. These methods find particular use in diagnosing and treating patients, e.g. infants, that are suspected of having NEC, IP, or sepsis.

Obtaining an expression signature.

[0061] In practicing methods of the invention, an expression signature, e.g. an NEC-Dx expression signature, a sepsis-Dx expression signature, or an NEC-M/S expression signature, is obtained for a patient that is suspected of having NEC or sepsis. An "expression signature" is a representation of the expression levels of one or more genes of interest in a patient sample, and comprises quantitative data on the expression levels of those one or more genes of interest in that sample. For example, an NEC-Dx signature is a representation of the expression levels of one or more NEC-Dx genes, where an NEC-Dx gene is a gene that is upregulated, i.e. is expressed at a higher level, in NEC patients relative to unaffected

individuals, e.g. healthy individuals, individuals with sepsis, etc. A sepsis-Dx signature is a representation of the expression levels of one or more sepsis-Dx genes, where a sepsis-Dx gene is a gene that is upregulated, i.e. is expressed at a higher level, in individuals having sepsis versus individuals that are not affected with sepsis, e.g. healthy individuals, individuals with NEC, etc. An NEC-M/S signature is a representation of the expression levels of one or more NEC-M and/or NEC-S genes, where an NEC-M gene is a gene that is expressed at a higher level, i.e. is upregulated, in M-class NEC patients relative to S-class NEC patients, and an NEC-S gene is a gene that is expressed at a higher level, i.e. is upregulated, in S-class NEC patients than in M-class patients. Non-limiting examples of NEC-Dx genes, sepsis-Dx genes, NEC-S, and NEC-M genes are provided in the table below.

Table 1. NEC-Dx genes, sepsis-Dx genes, NEC-M genes, and NEC-S genes. Sequences for genes are provided as Genbank Accession Entries, the disclosures of which are specifically incorporated herein by reference.

Class	Gene	Gene name, aliases	Genbank Accession No.
NEC-Dx	CD14	CD14 molecule	NM_000591.3 (variant 1) NM_001040021.2 (variant 2) NM_001174104.1 (variant 3) NM_001174105.1 (variant 4)
	SAP1	SH2 domain containing 1A; SAP; SH2D1A	NM_002351.4 (isoform 1); NM_001114937.2 (isoform 2)
	PEDF	serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1; SERPINF1	NM_002615.4
	Q6ZUQ4	CDNA FLJ43449 fis	Q6ZUQ4 (protein database)
	OBFC2B	oligonucleotide/oligosaccharide-binding fold containing 2B	NM_024068.3
	COL11A2	collagen, type XI, alpha 2	NM_080680.2 (isoform 1), NM_080681.2 (isoform 2), NM_080679.2 (isoform 3), NM_001163771 (isoform 4)
	NBEAL2	neurobeachin-like 2	NM_015175.1
	GRASP	GRP1 (general receptor for phosphoinositides 1)-associated scaffold protein	NMJ8171 1.2
	HUWE1	HECT, UBA and WWE domain containing 1	NM_031407.4
	COL1A2	collagen, type I, alpha 2	NM_000089.3
	HOXD3	homeobox D3	NM_006898.4
	DSG4	desmoglein 4	NM_001134453.1 (variant 1) NM_177986.3 (variant 2)
	KRTAP5-11	keratin associated protein 5-11	NM_001005405.2

	Y1020	hypothetical protein Y1020 [Yersinia pestis KIM].	NP_857803
	FGA	fibrinogen alpha chain	NM_021871.2 (isoform α) NM_000508.3 (isoform α-E)
	CTAPIII	pro-platelet basic protein (chemokine (C-X-C motif) ligand 7; PPBP	NM_002704.3
	SAA1	serum amyloid A1	NM_000331.4
	B2M	beta-2-microglobulin	NM_004048.2
	TTR	Transthyretin	NM_000371.3
	OSTP	Osteopontin; OPN; secreted phosphoprotein 1, SPP1, BNSP; BSPI; ETA-1; MGC110940	NM_001040058.1 NM_001040058.1 NM_000582.2
	APOA4	apolipoprotein A-IV	NM_000482.3
	C08G	Complement component C8 gamma chain; C8G	NM_000606.2
	ANGT	Angiotensinogen; serpin peptidase inhibitor, clade A, member 8; AGT	NM_000029.3
	FIBA	Fibrinogen alpha chain; FGA	NM_000508.3 NM_021871.2
	PROF1	Profilin 1; PFN1	NM_005022.2
	UMOD	Uromodulin	NM_003361.2 (variant 1) NM_001008389.1 (variant 2)
	PLSL	Plastin-2; lymphocyte cytosolic protein 1; LCP1	NM_002298.4
	LMAN2	lectin, mannose-binding 2	NM_006816.2
Sepsis-Dx	ftsY	ECK3448, b3464, JW3429 Bacterial signal recognition particle receptor	NP_417921.1
	PROC	protein C (inactivator of coagulation factors Va and VIIIa)	NM_000312.3
	MAP1B	microtubule-associated protein 1B	NM_005909.3
	CSN5	COP9 constitutive photomorphogenic homolog subunit 5 (Arabidopsis); COPS5	NM_006837.2
NEC-S	Q6ZUQ4	CDNA FLJ43449 fis	Q6ZUQ4 (protein database)
	OBFC2B	oligonucleotide/oligosaccharide-binding fold containing 2B	NM_024068.3
	COL11A2	collagen, type XI, alpha 2	NM_080680.2 (isoform 1), NM_080681.2 (isoform 2), NM_080679.2 (isoform 3), NM_001163771 (isoform 4)
	NBEAL2	neurobeachin-like 2	NM_015175.1
	GRASP	GRP1 (general receptor	NM_181711.2

		for phosphoinositides 1)-associated scaffold protein	
	HUWE1	HECT, UBA and WWE domain containing 1	NM_031407.4
	COL1A2	collagen, type I, alpha 2	NM_000089.3
	HOXD3	homeobox D3	NM_006898.4
	DSG4	desmoglein 4	NM_001134453.1 (variant 1) NM_177986.3 (variant 2)
	KRTAP5-11	keratin associated protein 5-11	NM_001005405.2
	Y1020	hypothetical protein Y1020 [Yersinia pestis KIM].	NP_857803
	FGA	fibrinogen alpha chain	NM_021871.2 (isoform α) NM_000508.3 (isoform α-E)
	CTAPIII	pro-platelet basic protein (chemokine (C-X-C motif) ligand 7; PPBP	NM_002704.3
	SAA1	serum amyloid A1	NM_000331.4
	B2M	beta-2-microglobulin	NM_004048.2
	TTR	Transthyretin	NM_000371.3
	OSTP	Osteopontin; OPN; secreted phosphoprotein 1, SPP1, BNSP; BSPI; ETA-1; MGC110940	NM_001040058.1 NM_000582.2
	APOA4	apolipoprotein A-IV	NM_000482.3
	C08G	Complement component C8 gamma chain; C8G	NM_000606.2
	SAP1	SH2 domain containing 1A; SAP; SH2D1A	NM_002351.4 (isoform 1); NM_001114937.2 (isoform 2)
	ANGT	Angiotensinogen; serpin peptidase inhibitor, clade A, member 8; AGT	NM_000029.3
	CD14	CD14 molecule	NM_000591.3 (variant 1) NM_001040021.2 (variant 2) NM_001174104.1 (variant 3) NM_001174105.1 (variant 4)
	FIBA	Fibrinogen alpha chain; FGA	NM_000508.3 NM_021871.2
	PROF1	Profilin 1; PFN1	NM_005022.2
	PEDF	serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1; SERPINF1	NM_002615.4
NEC-M	UMOD	Uromodulin	NM_003361.2 (variant 1) NM_001008389.1 (variant 2)
	PLSL	Plastin-2;; lymphocyte cytosolic protein 1; LCP1	NM_002298.4
	LMAN2	lectin, mannose-binding 2	NM_006816.2

[0062] In practicing methods of the invention, an expression signature, e.g. a NEC-Dx expression signature, a sepsis-Dx expression signature, or an NEC-M/S expression signature, is obtained for a patient. In some embodiments, the patient is suspected of having NEC or sepsis. A patient that is suspected of having NEC or sepsis is one in which historical factors, physical findings and radiological findings that indicate risk for NEC or sepsis. Historical factors include, for example, feeding intolerance (defined as vomiting two or more feedings within 24 hours or any vomit containing bile, or the presence of gastric residuals of volume greater than 6 cc/kg or any aspirate containing bile), apneic/bradycardic episodes, oxygen desaturation episodes, guaiac positive, or bloody stools. Physical findings include, for example, abdominal distention, capillary refill time > 2sec, abdominal wall discoloration, or abdominal tenderness. Radiological findings include, for example, pneumatosis intestinalis, portal venous gas, Ileus, dilated bowel, pneumoperitoneum, air/fluid levels, thickened bowel walls, ascites or peritoneal fluid, or free intraperitoneal air, absent bowel sounds, hypotension, abdominal cellulitis, and right lower quadrant mass.

[0063] To obtain an expression signature, the expression level of the one or more genes of interest is measured, *i.e.* the expression levels of 1 or more, 2 or more, or 3 or more genes is determined, e.g. 4 or more, 5 or more, 6 or more or 7 or more genes, in some embodiments8-15 genes, in some embodiments 16-28 genes, e.g. the expression levels of 28 or more genes is determined. The expression level is typically measured by analyzing a body fluid sample, e.g. a sample of urine, blood, or saliva, that is obtained from an individual. The sample that is collected may be freshly assayed or it may be stored and assayed at a later time. If the latter, the sample may be stored by any convenient means that will preserve the sample so that gene expression may be assayed at a later date. For example the sample may freshly cryopreserved, that is, cryopreserved without impregnation with fixative, e.g. at 4°C, at -20°C, at -60°C, at -80°C, or under liquid nitrogen. Alternatively, the sample may be fixed and preserved, e.g. at room temperature, at 4°C, at -20°C, at -60°C, at -80°C, or under liquid nitrogen, using any of a number of fixatives known in the art, e.g. alcohol, methanol, acetone, formalin, paraformaldehyde, etc.

[0064] The sample may be assayed as a whole sample, *e.g.* in crude form. Alternatively, the sample may be fractionated prior to analysis, e.g. for a blood sample, to purify leukocytes if, e.g., the gene expression product to be assayed is RNA or intracellular protein, or to purify plasma or serum if, e.g., the gene expression product is a secreted polypeptide. Further fractionation may also be performed, e.g., for a purified leukocyte sample, fractionation by e.g. panning, magnetic bead sorting, or fluorescence activated cell sorting (FACS) may be performed to enrich for particular types of cells, thereby arriving at an enriched population of that cell type for analysis; or, e.g., for a plasma or serum sample, fractionation based upon size, charge, mass, or other physical characteristic may be performed to purify particular

secreted polypeptides, e.g. under denaturing or non-denaturing ("native") conditions, depending on whether or not a non-denatured form is required for detection. One or more fractions are then assayed to measure the expression levels of the one or more genes of interest.

[0065] The expression levels of the one or more genes of interest may be measured by polynucleotide, i.e. mRNA, levels, or by protein level. Any convenient method for measuring mRNA levels in a sample may be used, e.g. hybridization-based methods, e.g. northern blotting and in situ hybridization (Parker & Barnes, Methods in Molecular Biology 106:247-283 (1999)), RNAse protection assays (Hod, Biotechniques 13:852-854 (1992)), and PCR-based methods (e.g. reverse transcription PCR (RT-PCR) (Weis et al., Trends in Genetics 8:263-264 (1992)). Alternatively, any convenient method for measuring protein levels in a sample may be used, e.g. antibody-based methods, e.g. immunoassays, e.g., enzyme-linked immunosorbent assays (ELISAs), immunohistochemistry, and flow cytometry (FACS).

[0066] For measuring mRNA levels, the starting material may betotal RNA, i.e. unfractionated RNA, or poly A+ RNA isolated from a suspension of cells, e.g. a peripheral blood sample. General methods for mRNA extraction are well known in the art and are disclosed in standard textbooks of molecular biology, including Ausubel et al., Current Protocols of Molecular Biology, John Wiley and Sons (1997). RNA isolation can also be performed using a purification kit, buffer set and protease from commercial manufacturers, according to the manufacturer's instructions. For example, RNA from cell suspensions can be isolated using Qiagen RNeasy mini-columns, and RNA from cell suspensions or homogenized tissue samples can be isolated using the TRIzol reagent-based kits (Invitrogen), MasterPure™ Complete DNA and RNA Purification Kit (EPICENTRE™, Madison, WI), Paraffin Block RNA Isolation Kit (Ambion, Inc.) or RNA Stat-60 kit (Tel-Test).

[0067] mRNA levels may be measured by any convenient method. Examples of methods for measuring mRNA levels may be found in, e.g., the field of differential gene expression analysis. One representative and convenient type of protocol for measuring mRNA levels is array-based gene expression profiling. Such protocols are hybridization assays in which a nucleic acid that displays "probe" nucleic acids for each of the genes to be assayed/profiled in the profile to be generated is employed. In these assays, a sample of target nucleic acids is first prepared from the initial nucleic acid sample being assayed, where preparation may include labeling of the target nucleic acids with a label, e.g., a member of signal producing system. Following target nucleic acid sample preparation, the sample is contacted with the array under hybridization conditions, whereby complexes are formed between target nucleic acids that are complementary to probe sequences attached to the array surface. The presence of hybridized complexes is then detected, either qualitatively or quantitatively.

[0068] Specific hybridization technology which may be practiced to generate the expression profiles employed in the subject methods includes the technology described in U.S. Patent Nos.: 5,143,854; 5,288,644; 5,324,633; 5,432,049; 5,470,710; 5,492,806; 5,503,980; 5,510,270; 5,525,464; 5,547,839; 5,580,732; 5,661,028; 5,800,992; the disclosures of which are herein incorporated by reference; as well as WO 95/21265; WO 96/31622; WO 97/10365; WO 97/27317; EP 373 203; and EP 785 280. In these methods, an array of "probe" nucleic acids that includes a probe for each of the phenotype determinative genes whose expression is being assayed is contacted with target nucleic acids as described above. Contact is carried out under hybridization conditions, e.g., stringent hybridization conditions, and unbound nucleic acid is then removed. The term "stringent assay conditions" as used herein refers to conditions that are compatible to produce binding pairs of nucleic acids, e.g., surface bound and solution phase nucleic acids, of sufficient complementarity to provide for the desired level of specificity in the assay while being less compatible to the formation of binding pairs between binding members of insufficient complementarity to provide for the desired specificity. Stringent assay conditions are the summation or combination (totality) of both hybridization and wash conditions.

[0069] The resultant pattern of hybridized nucleic acid provides information regarding expression for each of the genes that have been probed, where the expression information is in terms of whether or not the gene is expressed and, typically, at what level, where the expression data, i.e., expression profile (e.g., in the form of a transcriptosome), may be both qualitative and quantitative.

[0070] Additionally or alternatively, non-array based methods for quantitating the level of one or more nucleic acids in a sample may be employed. These include those based on amplification protocols, e.g., Polymerase Chain Reaction (PCR)-based assays, including quantitative PCR, reverse-transcription PCR (RT-PCR), real-time PCR, and the like, e.g. TaqMan® RT-PCR, MassARRAY® System, BeadArray® technology, and Luminex technology; and those that rely upon hybridization of probes to filters, e.g. Northern blotting and in situ hybridization.

[0071] For measuring protein levels, the amount or level in the sample of one or more proteins/polypeptides encoded by the gene of interest is determined. In such cases, any convenient protocol for evaluating protein levels may be employed wherein the level of one or more proteins in the assayed sample is determined.

[0072] While a variety of different manners of assaying for protein levels are known in the art, one representative and convenient type of protocol for assaying protein levels is ELISA. In ELISA and ELISA-based assays, one or more antibodies specific for the proteins of interest may be immobilized onto a selected solid surface, preferably a surface exhibiting a protein affinity such as the wells of a polystyrene microtiter plate. After washing to remove

incompletely adsorbed material, the assay plate wells are coated with a non-specific "blocking" protein that is known to be antigenically neutral with regard to the test sample such as bovine serum albumin (BSA), casein or solutions of powdered milk. This allows for blocking of non-specific adsorption sites on the immobilizing surface, thereby reducing the background caused by non-specific binding of antigen onto the surface. After washing to remove unbound blocking protein, the immobilizing surface is contacted with the sample to be tested under conditions that are conducive to immune complex (antigen/antibody) formation. Such conditions include diluting the sample with diluents such as BSA or bovine gamma globulin (BGG) in phosphate buffered saline (PBS)/Tween or PBS/Triton-X 100, which also tend to assist in the reduction of nonspecific background, and allowing the sample to incubate for about 2-4 hrs at temperatures on the order of about 25o-27oC (although other temperatures may be used). Following incubation, the antisera-contacted surface is washed so as to remove non-immunocomplexed material. An exemplary washing procedure includes washing with a solution such as PBS/Tween, PBS/Triton-X 100, or borate buffer. The occurrence and amount of immunocomplex formation may then be determined by subjecting the bound immunocomplexes to a second antibody having specificity for the target that differs from the first antibody and detecting binding of the second antibody. In certain embodiments, the second antibody will have an associated enzyme, e.g. urease, peroxidase, or alkaline phosphatase, which will generate a color precipitate upon incubating with an appropriate chromogenic substrate. For example, a urease or peroxidase-conjugated anti-human IgG may be employed, for a period of time and under conditions which favor the development of immunocomplex formation (e.g., incubation for 2 hr at room temperature in a PBS-containing solution such as PBS/Tween). After such incubation with the second antibody and washing to remove unbound material, the amount of label is quantified, for example by incubation with a chromogenic substrate such as urea and bromocresol purple in the case of a urease label or 2,2'-azino-di-(3-ethyl-benzthiazoline)-6-sulfonic acid (ABTS) and H202, in the case of a peroxidase label. Quantitation is then achieved by measuring the degree of color generation, e.g., using a visible spectrum spectrophotometer.

[0073] The preceding format may be altered by first binding the sample to the assay plate. Then, primary antibody is incubated with the assay plate, followed by detecting of bound primary antibody using a labeled second antibody with specificity for the primary antibody.

[0074] The solid substrate upon which the antibody or antibodies are immobilized can be made of a wide variety of materials and in a wide variety of shapes, e.g., microtiter plate, microbead, dipstick, resin particle, etc. The substrate may be chosen to maximize signal to noise ratios, to minimize background binding, as well as for ease of separation and cost. Washes may be effected in a manner most appropriate for the substrate being used, for example, by removing a bead or dipstick from a reservoir, emptying or diluting a reservoir such

as a microtiter plate well, or rinsing a bead, particle, chromatograpic column or filter with a wash solution or solvent.

[0075] Alternatively, non-ELISA based-methods for measuring the levels of one or more proteins in a sample may be employed. Representative examples include but are not limited to mass spectrometry, proteomic arrays, xMAPTM microsphere technology, western blotting, immunohistochemistry, flow cytometry, and detection in body fluid by electrochemical sensor. In, for example, flow cytometry methods, the quantitative level of gene products of the one or more genes of interest are detected on cells in a cell suspension by lasers. As with ELISAs and immunohistochemistry, antibodies (e.g., monoclonal antibodies) that specifically bind the polypeptides encoded by the genes of interest are used in such methods. As another example, electrochemical sensors may be employed. In such methods, a capture aptamer or an antibody that is specific for a target protein (the "analyte") is immobilized on an electrode. A second aptamer or antibody, also specific for the target protein, is labeled with, for example, pyrroquinoline quinone glucose dehydrogenase ((PQQ)GDH). The sample of body fluid is introduced to the sensor either by submerging the electrodes in body fluid or by adding the sample fluid to a sample chamber, and the analyte allowed to interact with the labeled aptamer/antibody and the immobilized capture aptamer/antibody. Glucose is then provided to the sample, and the electric current generated by (PQQ)GDH is observed, where the amount of electric current passing through the electrochemical cell is directly related to the amount of analyte captured at the electode.

[0076] The resultant data provides information regarding expression for each of the genes that have been probed, wherein the expression information is in terms of whether or not the gene is expressed and, typically, at what level, and wherein the expression data may be both qualitative and quantitative.

[0077] Once the expression level of the one or more genes of interest, e.g. NEC-Dx genes, sepsis Dx genes, NEC-M genes, NEC-S genes, has been determined, the measurement(s) may be analyzed in any of a number of ways to obtain an expression signature.

[0078] For example, an expression signature may be obtained by analyzing the data to generate an expression profile. As used herein, an expression profile is the normalized expression level of one or more genes of interest in a patient sample. An expression profile may be generated by any of a number of methods known in the art. For example, the expression level of each gene may be $\log_2$ transformed and normalized relative to the expression of a selected housekeeping gene, e.g. ABL1 , GAPDH, or PGK1 , or relative to the signal across a whole microarray, etc. An expression profile is one example of an expression signature.

[0079] As another example, an expression signature may be obtained by analyzed the data to generate an expression score. An expression score is a single metric value that represents

the sum of the weighted expression levels of one or more genes of interest in a patient sample. An expression score for a patient sample may be calculated by any of a number of methods known in the art for calculating gene signatures. For example, the expression levels of each of the one or more genes of interest in a patient sample may be $\log_2$ transformed and normalized, e.g. as described above for generating an expression profile. The normalized expression levels for each gene is then weighted by multiplying the normalized level to a weighting factor, or "weight", to arrive at weighted expression levels for each of the one or more genes, where the weights are defined by a reference dataset, or "training dataset", e.g. by Principle Component Analysis, Linear discriminant analysis (LDA), Fisher's linear discriminant analysis, etc, of a reference dataset. The weighted expression levels are then totaled and in some cases averaged to arrive at a single weighted expression level for the one or more genes analyzed. Any dataset relating to patients having NEC may be used as a reference dataset. For example, the weights may be determined based upon any of the datasets provided in the examples section below. Thus, the NEC-Dx score sepsis-Dx score, or NEC-M/S score is the first principle component of the NEC-Dx genes, the sepsis-Dx genes, or the NEC-M/S genes, respectively, in a sample as defined by a reference dataset.

[0080] As discussed above, expression signatures are obtained by analyzing data on expression levels to arrive at an expression profile or an expression score. This analysis may be readily performed by one of ordinary skill in the art by employing a computer-based system, e.g. using any hardware, software and data storage medium as is known in the art, and employing any algorithms convenient for such analysis.

Employing an NEC-Dx expression signature a Sepsis-Dx expression signature, or an NEC-M/S expression signature to evaluate a subject.

[0081] The NEC-Dx expression signature, sepsis-Dx expression signature, or NEC-M/S expression signature that is obtained may be employed to diagnose a NEC or sepsis, to provide a prognosis to a patient with NEC, or to provide a prediction of the responsiveness of a patient with NEC to a medical therapy. Typically, an expression signature is employed by comparing the expression signature to a reference or control, and using the results of that comparison (a "comparison result") to determine a diagnosis, prognosis or prediction. The terms "reference" and "control" as used herein mean a standardized gene expression profile, gene signature, or gene score to be used to interpret the NEC-Dx expression signature, Sepsis-Dx expression signature, or NEC-M/S expression signature of a given patient and assign a diagnostic, prognostic, and/or responsiveness class thereto. The reference or control is typically an expression profile or expression score that is obtained from a body fluid or tissue with a known association with a particular phenotype. Additionally, if the expression signature is an expression profile, the reference will typically be an expression profile from a control

sample, whereas if the expression signature is an expression score, the reference will typically be the expression score from a control sample.

[0082] For example, as disclosed in greater detail in the examples section below, high-risk phenotypes, e.g. increased expression of particular panels of genes, are associated with samples from certain patient cohorts, i.e. positive controls, e.g. increased expression of NEC-Dx genes, sepsis-Dx genes, NEC-M genes, or NEC-S genes is associated with samples from patients with NEC, with sepsis, with M-class NEC, or with S-class NEC, respectively. Thus, a positive control reference that may be used when making an NEC diagnosis could be a NEC-Dx signature (e.g. NEC-Dx profile or NEC-Dx score) of a body fluid sample from a patient with NEC; the positive control reference when making a sepsis diagnosis could be a sepsis-Dx signature (e.g. sepsis-Dx profile or sepsis-Dx score) of a body fluid sample from a patient with sepsis; and the positive control reference when providing a prognosis for an individual with NEC or predicting responsiveness of an individual with NEC to medical therapy may be an NEC-M/S signature (e.g. NEC-M/S profile or NEC-M/S score) of a body fluid sample from a patient with either M-class NEC or with S-class NEC.

[0083] As another example, low-risk phenotypes e.g. normal expression of particular panels of genes, are associated with sample from unaffected patients, i.e., negative controls. Thus, the negative control reference when making an NEC diagnosis may be aNEC-Dx signature (e.g. NEC-Dx profile or NEC-Dx score) of a body fluid sample from an individual that is not affected with NEC, e.g. a healthy individual or an individual with sepsis. Likewise, The negative control reference when making a sepsis diagnosis may be a sepsis-Dx signature (e.g. sepsis-Dx profile or sepsis-Dx score) of a body fluid sample from an individual that is not affected with sepsis, e.g. a healthy individual, or an individual with NEC. Similarly, The negative control reference when providing an M-class NEC prognosis may be a NEC-M signature (e.g. NEC-M profile or NEC-M score) of a body fluid sample from an individual that is not affected with M-class NEC, where a higher expression signature for M-genes in the patient sample than in the negative control (reference indicates that the patient has a high risk of M-class NEC. and a low risk of S-class NEC and, reciprocally, an expression signature for M-genes in the patient sample that is comparable to or lower than the expression signature in the negative control reference indicates that the patient has a low risk of M-class NEC and a high risk of S-class NEC. The negative control reference when providing an S-class NEC prognosis may be an NEC-S signature (e.g. NEC-S profile or NEC-S score) of a body fluid sample from an individual that is not affected with S-class NEC, where a higher expression signature of S-genes in the patient sample than in the negative control reference indicates that the patient has a high risk of S-class NEC and a low risk of M-class NEC and, reciprocally, an expression signature of S-genes in the patient sample that is comparable to or lower than the expression signature in the

negative control reference indicates that the patient has a low risk of S-class NEC and a high risk of M-class NEC.

[0084] In certain embodiments, the obtained expression signature is compared to a single reference/control expression signature to obtain information regarding the phenotype of the tissue being assayed. In yet other embodiments, the obtained expression signature is compared to two or more different reference/control expression signature to obtain more in-depth information regarding the phenotype of the assayed tissue. For example, an expression profile may be compared to both a positive expression profile and a negative expression profile, or an expression score may be compared to both a positive expression score and a negative expression score to obtain confirmed information regarding whether the tissue has the phenotype of interest. As another example, an expression profile or score may be compared to multiple expression profiles or scores, each correlating with a particular diagnosis, prognosis or therapeutic responsiveness, e.g. as might be provided in a report or table that discloses the correlation between particular NEC-Dx, sepsis-Dx, or NEC-M/S signatures and particular disease diagnoses, disease prognoses, or responsiveness to therapy.

[0085] As discussed above, an NEC-Dx signature may be employed to make an NEC diagnosis. For example, a patient can be diagnosed as being at high risk for having NEC or as being at low risk for having NEC depending on whether his NEC-Dx signature correlates more closely with the median NEC-Dx signature across a cohort of patients with NEC or whether his signature correlates more closely with the median NEC-Dx signature across a cohort of individuals unaffected by NEC. By "correlates closely", it is meant is within about 40% of the reference signature, e.g. 40%, 35%, or 30%, in some embodiments within 25%, 20%, or 15%, sometimes within 10%, 8%, 5%, or less. Alternatively, when two or more references are used, e.g. both a reference from a cohort of patient with NEC and a reference from a cohort of unaffected individuals, a patient can be diagnosed as being at high risk for having NEC or as being at low risk for having NEC depending on whether his signature correlates more closely with the median NEC-Dx signature across a cohort of patients with NEC or a cohort of individual unaffected by NEC.

[0086] Similarly, a sepsis-Dx signature may be employed to make a sepsis diagnosis. For example, a patient can be diagnosed as being at high risk for having sepsis or as being at low risk for having sepsis depending on whether his sepsis-Dx signature correlates more closely with the median sepsis-Dx signature across a cohort of patients with sepsis or whether his signature correlates more closely with the median sepsis-Dx signature across a cohort of individuals unaffected by NEC. As another example, a patient can be diagnosed as being at high risk for having sepsis or as being at low risk for having sepsis depending on whether his

sepsis-Dx signature correlates more closely with the median sepsis-Dx signature across a cohort of patients with sepsis or a cohort of individuals unaffected by sepsis.

[0087] In some embodiments, both an NEC diagnosis and a sepsis diagnosis may be made at the same time. In such embodiments, the gene expression levels of one or more of the NEC-Dx genes is measured at the same time that gene expression levels of one or more of the sepsis-Dx genes is measured. In certain embodiments, the NEC-Dx signature and the sepsis-Dx signature may be compared individually, i.e. separately, to signatures from one or more reference sample, i.e. the NEC-Dx signature is compared to an NEC-Dx signature from a reference sample, and the sepsis-Dx signature is compared to a sepsis-Dx signature from a reference sample, e.g. the same reference sample, and the results of the comparisons are employed to provide a prognosis for the patient. For example, a patient can be diagnosed as being at high risk for having NEC and at low risk for having sepsis or as being at low risk for having NEC and at high risk for having sepsis depending on whether his NEC-Dx and sepsis-Dx signatures correlate more closely with the median NEC-Dx and sepsis-Dx signatures across a cohort of individuals that have NEC, or more closely with the median NEC-Dx and sepsis-Dx signatures across a cohort of individuals that have sepsis. In certain embodiments, the NEC-Dx signature and the sepsis signature are combined to arrive at an NEC/sepsis-Dx signature, the NEC/sepsis-Dx signature is compared to a NEC/sepsis-Dx signature from a reference sample, and the results of the comparisons employed to provide a prognosis for the patient. For example, a patient can be diagnosed as being at high risk for having NEC and at low risk for having sepsis or as being at low risk for having NEC and at high risk for having sepsis depending on whether his combined NEC-Dx signature and sepsis-Dx signature (i.e. his NEC/sepsis-Dx signature) correlates more closely with the median combined NEC-Dx and sepsis-Dx signature across a cohort of patients that have NEC or a cohort of patients that have sepsis.

[0088] As also discussed above, an NEC-M/S expression signature may be employed to provide a prognosis to a patient suspected of or diagnosed as having NEC. For example, a patient can be ascribed to high- or low-risk categories, or high-, medium- or low- risk categories for overall survival depending on whether their NEC-M/S signature correlates more closely with the median NEC-M/S signature across a cohort of patients with the M class of the disease or the S class of the disease patient, the overall survival rates of patients with M class NEC or S class NEC being known in the art or readily determined by the ordinarily skilled artisans by, e.g., Kaplan-Meier analysis of individuals with M-class NEC and S-class NEC.

[0089] As also discussed above, an NEC-M/S expression signature may be employed to provide a prediction of responsiveness of a patient to a particular therapy, e.g. medical therapy. These predictive methods can be used to assist patients and physicians in making

treatment decisions, e.g. in choosing the most appropriate treatment modalities for any particular patient.

[0090] Additionally, the NEC-M/S expression signature may be used on samples collected from patients in a clinical trial and the results of the test used in conjunction with patient outcomes in order to determine whether subgroups of patients are more or less likely to show a response to a new drug than the whole group or other subgroups. Further, such methods can be used to identify from clinical data the subsets of patients who can benefit from therapy. Additionally, a patient is more likely to be included in a clinical trial if the results of the test indicate a higher likelihood that the patient will be responsive to medical treatment, and a patient is less likely to be included in a clinical trial if the results of the test indicate a lower likelihood that the patient will be responsive to medical treatment.

[0091] The subject methods can be used alone or in combination with other clinical methods for patient stratification known in the art to provide a diagnosis, a prognosis, or a prediction of responsiveness to therapy. For example, clinical parameters that are known in the art for diagnosing NEC, diagnosing types of NEC, or staging NEC, or for diagnosing or staging sepsis, may also be incorporated into the ordinarily skilled artisan's analysis to arrive at a diagnosis, prognosis, or prediction of responsiveness to therapy with the subject methods.

[0092] For example, one common clinically used set of criteria for staging Necrotizing Enterocolitis is Modified Bell's criteria, described in detail in Table 2 below. In some embodiments, an NEC clinical score may be obtained, that NEC score comprising data on the clinical findings regarding the patient, for example on the pH value of blood; portal venous gas in x-ray; abdominal ileus in x-ray; the use of a vasopressor prior to diagnosis; abdominal distention; whether cranial ultrasound was done for ivh (intra-ventricular hemorrhage); vasopressor on diagnosis, i.e. the patient is receiving medications that support blood pressure, e.g. inotropes, chronotropes, alpha agonists and the like, e.g. dopamine; ventilation on diagnosis; whether any positive culture of bacteria or fungus was obtained from blood or urine within 5 days of diagnosis; the gestational age of the patient at birth; (and the patient's birth weight. The NEC clinical score is then used in conjunction with the expression signature to provide an NEC diagnosis with greater accuracy, specificity and sensitivity. For example, the NEC-Dx signature and the NEC clinical score are compared to an NEC-Dx signature and an NEC clinical score from a reference sample, and the results of both comparisons are employed to provide an NEC diagnosis to the patient; or the NEC-M/S signature and the NEC clinical score are compared to a NEC-M/S signature and an NEC clinical score from a reference sample, and the results of both comparisons are employed to provide a sepsis diagnosis to the patient. In some embodiments, the NEC clinical score is used alongside the expression signature of the subject methods. In other embodiments, the NEC clinical score is integrated with the expression score to obtain a single metric value that is representative of both the NEC

clinical score and the expression score, i.e. an NEC-gene/clinic score (an "NEC-G/C score"), e.g. an NEC-Dx G/C score, or an NEC-M/S G/C score, where that integrated score is compared to an integrated score for a reference sample, at the results of this comparison are employed to provide a prognosis to the patient or to predict the responsiveness of the patient to medical therapy.

[0093] **Table 2:** Modified Bell's criteria for staging Necrotizing Enterocolitis. "NPO" = nothing by mouth

Stage	Systemic signs	Abdominal signs	Radiographic signs	Treatment
IA Suspected	Temperature instability, apnea, bradycardia, lethargy	Gastric retention, abdominal distention, emesis, heme-positive stool	Normal or intestinal dilation, mild ileus	NPO, antibiotics x 3 days
IB Suspected	Same as above	Grossly bloody stool	Same as above	Same as IA
IIA Definite, mildly ill	Same as above	Same as above, plus absent bowel sounds with or without abdominal tenderness	Intestinal dilation, ileus, pneumatosis intestinalis	NPO, antibiotics x 7 to 10 days
IIB Definite, moderately ill	Same as above, plus mild metabolic acidosis and thrombocytopenia	Same as above, plus absent bowel sounds, definite tenderness, with or without abdominal cellulitis or right lower quadrant mass	Same as IIA, plus ascites	NPO, antibiotics x 14 days
IIIA Advanced, severely ill, intact bowel	Same as IIB, plus hypotension, bradycardia, severe apnea, combined respiratory and metabolic acidosis, Disseminated Intravascular Coagulation (DIC), and neutropenia	Same as above, plus signs of peritonitis, marked tenderness, and abdominal distention	Same as IIA, plus ascites	NPO, antibiotics x 14 days, fluid resuscitation, inotropic support, ventilator therapy, paracentesis

IIIB Advanced, severely ill, perforated bowel	Same as IIIA	Same as IMA	Same as above, plus pneumo-peritoneum	Same as IIA, plus surgery

[0094] As another example, the American College of Chest Physicians and the Society of Critical Care Medicine describes several different levels of sepsis (see Table 3, below). In some embodiments, a sepsis clinical score may be obtained, that sepsis clinical score comprising data on the clinical findings regarding the patient as described in the table. The sepsis clinical score is then used in conjunction with the expression signature to provide a sepsis diagnosis with greater accuracy, specificity and sensitivity. For example, the sepsis-Dx signature and the sepsis clinical score are compared to a sepsis-Dx signature and a sepsis clinical score from a reference sample, and the results of both comparisons are employed to provide an sepsis diagnosis to the patient. In some embodiments, the sepsis clinical score is used alongside the expression signature of the subject methods. In other embodiments, the sepsis clinical score is integrated with the expression score to obtain a single metric value that is representative of both the sepsis clinical score and the expression score, i.e. a sepsis-Dx G/C score, where that integrated score is compared to an integrated score for a reference sample, at the results of this comparison are employed to provide a prognosis to the patient or to predict the responsiveness of the patient to medical therapy.

[0095] **Table** 3: Sepsis levels, as described by the American College of Chest Physicians and the Society of Critical Care Medicine

* **Systemic inflammatory response syndrome (SIRS).**

  Defined by the presence of two or more of the following findings:

  o Body temperature < 36 °C (97 °F) or > 38 °C (100 °F) (hypothermia or fever).

  o Heart rate > 90 beats per minute.

  o Respiratory rate > 20 breaths per minute or, on blood gas, a PaC02 less than 32 mm Hg (4.3 kPa) (tachypnea or hypocapnia due to hyperventilation).

  o White blood cell count < 4,000 cells/mm3 or > 12,000 cells/mm3 (< 4 × 109 or > 12 × 109 cells/L), or greater than 10% band forms (immature white blood cells). (leukopenia, leukocytosis, or bandemia).

* **Sepsis.**

  Defined as SIRS in response to a confirmed infectious process. Infection can be suspected or proven (by culture, stain, or polymerase chain reaction (PCR)), or a clinical syndrome pathognomonic for infection. Specific evidence for infection includes WBCs in normally sterile fluid (such as urine or cerebrospinal fluid

(CSF), evidence of a perforated viscus (free air on abdominal x-ray or CT scan, signs of acute peritonitis), abnormal chest x-ray (CXR) consistent with pneumonia (with focal opacification), or petechiae, purpura, or purpura fulminans

* **Severe sepsis.**

Defined as sepsis with organ dysfunction, hypoperfusion, or hypotension.

* **Septic shock.**

Defined as sepsis with refractory arterial hypotension or hypoperfusion abnormalities in spite of adequate fluid resuscitation. Signs of systemic hypoperfusion may be either end-organ dysfunction or serum lactate greater than 4 mmol/dL. Other signs include oliguria and altered mental status. Patients are defined as having septic shock if they have sepsis plus hypotension after aggressive fluid resuscitation (typically upwards of 6 liters or 40 ml/kg of crystalloid).

[0096] In some embodiments, providing an evaluation of a subject with suspected or confirmed NEC or sepsis, i.e., a diagnosis of NEC or of sepsis, a prognosis for a patient with NEC, or a prediction of responsiveness of a patient with NEC to therapy, includes generating a written report that includes the artisan's assessment of the subject's current state of health i.e. a "diagnosis assessment", of the subject's prognosis, i.e. a "prognosis assessment", and/or of possible treatment regimens, i.e. a "treatment assessment". Thus, a subject method may further include a step of generating or outputting a report providing the results of a diagnosis assessment, a prognosis assessment, or treatment assessment, which report can be provided in the form of an electronic medium (e.g., an electronic display on a computer monitor), or in the form of a tangible medium (e.g., a report printed on paper or other tangible medium).

[0097] A "report," as described herein, is an electronic or tangible document which includes report elements that provide information of interest relating to a diagnosis assessment, a prognosis assessment, and/or a treatment assessment and its results. A subject report can be completely or partially electronically generated. A subject report includes at least a diagnosis assessment, i.e. a diagnosis as to whether a subject has a high likelihood of having NEC or sepsis; or a prognosis assessment, i.e. a prediction of the likelihood that a patient with NEC will have an NEC-attributable death; or a treatment assessment, i.e. a prediction as to the likelihood that an NEC patient will have a particular clinical response to treatment, and/or a suggested course of treatment to be followed. A subject report can further include one or more of: 1) information regarding the testing facility; 2) service provider information; 3) subject data; 4) sample data; 5) an assessment report, which can include various information including: a) test data, where test data can include i) the gene expression levels of one or more

NEC-Dx genes, sepsis-Dx genes, NEC-M genes or NEC-S genes, ii) the gene expression profiles for one or more NEC-Dx, sepsis-Dx, NEC-M or NEC-genes, and/or iii) an NEC-Dx, sepsis-Dx, or NEC-M/S score, b) reference values employed, if any; 6) other features.

[0098] The report may include information about the testing facility, which information is relevant to the hospital, clinic, or laboratory in which sample gathering and/or data generation was conducted. This information can include one or more details relating to, for example, the name and location of the testing facility, the identity of the lab technician who conducted the assay and/or who entered the input data, the date and time the assay was conducted and/or analyzed, the location where the sample and/or result data is stored, the lot number of the reagents (e.g., kit, etc.) used in the assay, and the like. Report fields with this information can generally be populated using information provided by the user.

[0099] The report may include information about the service provider, which may be located outside the healthcare facility at which the user is located, or within the healthcare facility. Examples of such information can include the name and location of the service provider, the name of the reviewer, and where necessary or desired the name of the individual who conducted sample gathering and/or data generation. Report fields with this information can generally be populated using data entered by the user, which can be selected from among pre-scripted selections (e.g., using a drop-down menu). Other service provider information in the report can include contact information for technical information about the result and/or about the interpretive report.

[00100] The report may include a subject data section, including subject medical history as well as administrative subject data (that is, data that are not essential to the diagnosis, prognosis, or treatment assessment) such as information to identify the subject (e.g., name, subject date of birth (DOB), gender, mailing and/or residence address, medical record number (MRN), room and/or bed number in a healthcare facility), insurance information, and the like), the name of the subject's physician or other health professional who ordered the susceptibility prediction and, if different from the ordering physician, the name of a staff physician who is responsible for the subject's care (e.g., primary care physician).

[00101] The report may include a sample data section, which may provide information about the biological sample analyzed, such as the source of biological sample obtained from the subject (e.g. blood, urine, saliva), how the sample was handled (e.g. storage temperature, preparatory protocols) and the date and time collected. Report fields with this information can generally be populated using data entered by the user, some of which may be provided as pre-scripted selections (e.g., using a drop-down menu).

[00102] The report may include an assessment report section, which may include information generated after processing of the data as described herein. The interpretive report can include a prognosis of the likelihood that the patient will have an NEC-attributable death or

progression. The interpretive report can include, for example, results of the gene expression analysis, methods used to calculate the NEC-Dx, sepsis-Dx, NEC-M/S signature, and interpretation, i.e. prognosis. The assessment portion of the report can optionally also include a Recommendation(s). For example, where the results indicate that the subject has NEC, the recommendation can include a recommendation that broad-spectrum antibiotics be provided and that no nutrition be provided by mouth.

[00103] It will also be readily appreciated that the reports can include additional elements or modified elements. For example, where electronic, the report can contain hyperlinks which point to internal or external databases which provide more detailed information about selected elements of the report. For example, the patient data element of the report can include a hyperlink to an electronic patient record, or a site for accessing such a patient record, which patient record is maintained in a confidential database. This latter embodiment may be of interest in an in-hospital system or in-clinic setting. When in electronic format, the report is recorded on a suitable physical medium, such as a computer readable medium, e.g., in a computer memory, zip drive, CD, DVD, etc.

[00104] It will be readily appreciated that the report can include all or some of the elements above, with the proviso that the report generally includes at least the elements sufficient to provide the analysis requested by the user (e.g., a diagnosis, a prognosis, or a prediction of responsiveness to a therapy).

REAGENTS, DEVICES AND KITS

[00105] Also provided are reagents, devices and kits thereof for practicing one or more of the above-described methods. The subject reagents, devices and kits thereof may vary greatly. Reagents and devices of interest include those mentioned above with respect to the methods of assaying gene expression levels, where such reagents may include RNA or protein purification reagents, nucleic acid primers specific for NEC-Dx genes, sepsis-Dx genes, NEC-M genes and/or NEC-S genes, arrays of nucleic acid probes, antibodies to NEC-Dx polypeptides, sepsis-Dx polypeptides, NEC-M polypeptides and/or NEC-S polypeptides (e.g., immobilized on a substrate), signal producing system reagents, etc., depending on the particular detection protocol to be performed. For example, reagents may include PCR primers that are specific for one or more of the NEC-Dx genes CD14, SAP1 , PEDF, Q6ZUQ4, OBFC2B, COL1 1A2, NBEAL2, GRASP, HUWE1 , COL1A2, HOXD3, DSG4, KRTAP5-1 1, Y1020, FGA, CTAPIII/PPBP, SAA1 , B2M, TTR, OSTP/OPN, APOA4, C08G, ANGT, FIBA, PROF1 , UMOD, PLSL, and LMAN2; PCR primers that are specific for one or more of the sepsis Dx genes ftsy, PROC, MAPI B and CSN5; PCR primers that are specific for the one or more NEC-M genes UMOD, PLSL, and LMAN2; and/or PCR primers that are specific for one or more of the NEC-S genes Q6ZUQ4, OBFC2B, COL1 1A2, NBEAL2, GRASP, HUWE1 ,

COL1A2, HOXD3, DSG4, KRTAP5-1 1, Y1020, FGA, OSTP/OPN, APOA4, C08G, SAP1 , ANGT, CD14, FIBA, PROF1 , and PEDF. Other examples of reagents include arrays that comprise probes that are specific for one or more of the NEC-Dx genes, sepsis-Dx genes, NEC-M genes and/or NEC-S genes; antibodies to epitopes of the proteins encoded by NEC-Dx genes, sepsis-Dx genes, NEC-M genes and/or NEC-S genes; or other reagents that may be used to detect the expression of NEC-Dx genes, sepsis-Dx genes, NEC-M genes and/or NEC-S genes.

[00106] The subject kits may also comprise one or more expression signature references, e.g. a reference for an NEC-Dx signature, a sepsis-Dx signature, and/or an NEC-M/S signature, for use in employing the expression signature obtained from a patient sample. For example, the reference may be a sample of a known phenotype, e.g. an unaffected individual, or an affected individual, e.g. from a particular risk group that can be assayed alongside the patient sample, or the reference may be a report of disease diagnosis, disease prognosis, or responsiveness to therapy that is known to correlate with one or more of the subject expression signatures.

[00107] In addition to the above components, the subject kits may further include instructions for practicing the subject methods. These instructions may be present in the subject kits in a variety of forms, one or more of which may be present in the kit. One form in which these instructions may be present is as printed information on a suitable medium or substrate, e.g., a piece or pieces of paper on which the information is printed, in the packaging of the kit, in a package insert, etc. Yet another means would be a computer readable medium, e.g., diskette, CD, DVD, etc., on which the information has been recorded. Yet another means that may be present is a website address which may be used via the internet to access the information at a removed site. Any convenient means may be present in the kits.

EXAMPLES

[00108] The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the present invention, and are not intended to limit the scope of what the inventors regard as their invention nor are they intended to represent that the experiments below are all or the only experiments performed. Efforts have been made to ensure accuracy with respect to numbers used (e.g. amounts, temperature, etc.) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is weight average molecular weight, temperature is in degrees Centigrade, and pressure is at or near atmospheric.

Example 1

[00109] Necrotizing enterocolitis (NEC) is a major cause of overall neonatal morbidity and mortality. Disease outcome for infants with NEC is largely determined by the degree of clinical

progression. Generally, half of affected infants recover with medical therapy alone (NEC M=medical class) and 30-50% develop progressive disease requiring surgery or resulting in death (NEC S=surgical class). Most of the disease associated morbidity, and nearly all of the mortality, occurs in the cohort with progressive disease requiring surgery. Previous attempts to identify clinical parameters that could reliably identify infants with NEC most likely to progress to severe disease have been unsuccessful. We hypothesized that an integrative analysis of clinical parameters along with protein biomarkers would result in a predictive algorithm of NEC progression. A multivariate analysis of patients (NEC 43 M and 26 S) using the standard NEC classification scheme of Bell failed to differentiate NEC outcomes. A novel panel of eleven clinical parameters, selected by Mann Whitney U test, (NEC 43 M and 26 S subjects) as a biomarker panel did stratify NEC subjects into low, intermediate and high-risk groups for progression. Molecular profiling of the urine peptidome (NEC 17 M and 11 S subjects) and plasma proteome (NEC 60 M and 30 S subjects) identified separate candidate biomarker panels of 36 urine peptides and 30 plasma proteins as biomarkers for progressive NEC. Complete clinical and molecular records were available for 13 NEC M and 11 NEC S patients affording detailed comparative analyses of the statistical performance of the clinical ($P=0.64$), urine ($P=9.5X10^{-4}$), and plasma panels ($P=1.3 \times 10^{-3}$) for NEC progression classification. Integrative analyses combining the clinical parameters, urine peptides and plasma proteins improved the NEC progression predictive performance (P value of 5.2×10^{-4}), leading to an optimal biomarker panel (15 urine peptides and 3 plasma proteins) that discriminates NEC M and S class with high sensitivity and specificity (P value of 4.0×10^{-7} and ROC AUC 0.99). We conclude that ensemble data mining methods utilizing clinical and molecular based classifiers produces effective predictive integrated algorithm for NEC progression.

MATERIALS AND METHODS

[00110] Clinical data collection. All 50 clinical and demographic parameters, summarized in Table 3, relevant to the initial diagnosis of NEC were extracted from an observational, prospective cohort study conducted by the NEC consortium consisting of the following institutions: Texas Children's Hospital, Houston; Lucile Packard Children's Hospital, Stanford; Johns Hopkins Children's Hospital, Baltimore; The Children's Hospital of Philadelphia; and Yale-New Haven Children's Hospital. In this study, infants who met at least one criterion from each of the three Modified Bell's criteria categories including historical factors, physical findings and radiological findings, were identified as suspicious for or diagnostic of NEC and became eligible for the study. Historical factors include feeding intolerance (defined as vomiting two or more feedings within 24 hours or any vomitus containing bile, or the presence of gastric residuals of volume greater than 6 cc/kg or any aspirate containing bile), apneic/bradycardic episodes, oxygen desaturation episodes, guaiac positive, or bloody stools.

Physical findings include abdominal distention, capillary refill time > 2sec, abdominal wall discoloration, or abdominal tenderness. Radiological findings include pneumatosis intestinalis, portal venous gas, Ileus, dilated bowel, pneumoperitoneum, air/fluid levels, thickened bowel walls, ascites or peritoneal fluid, or free intraperitoneal air. The following variables were not included in the consortium NEC database but are part of the Modified Bells Criteria: temperature instability, absent bowel sounds, hypotension, abdominal cellulitis, and right lower quadrant mass. A total of 15 clinical parameters were utilized, upon data availability, as the NEC modified staging criteria and are detailed in Table 4 and Figure 1A.

[00111] **Table** 4. Clinical parameters for NEC progression analysis

Clinical variables	Description of the clinical variables	Bells Criteria Used in initial screening for eligibility in the NEC database	Modified Bells Criteria	11 Clinical Parameters used in NEC stratification
patgen	Gender; 1=male 2=female			
datebirth	Date of birth			
gestage	gestational age			X
birthwt	birthweight in grams			X
prodef	date in which patient met definition of NEC			
TIMEBTWN	time between enrollment and specimen collection			
first_endpt	First endpoint: S=surg. T=transp. E=end fu X=non-nec F|R=full feeds D=death			
severedate	Date of 1st surgery or death			
surg1day	Day between enrollment and 1st surgery			
surg2day	Day between enrollment and 2nd surgery			
surgdate1	Date of 1st surgery			
surgdate2	Date of 2nd surgery			

deathdt	Date of death			
bloodstool	gross blood in stool; 1=yes 0=no	X	X	
abddistend	abdominal distention	X	X	X
caprefill	capilary refill time > 2 seconds	X	X	
abdcolor	abdominal discoloration	X	X	
abdpain	abdominal tenderness	X	X	
pintestin	pneumatosis intestinalis in x-ray	X	X	
portvenous	portal venous gas in x-ray	X	X	X
abdileus	abdominal ileus in x-ray	X	X	X
dilatebowel	dilated bowel in x-ray	X	X	
air	air or fluid in x-ray	X	X	
pnemo	pneumoperitoneum in x-ray	X	X	
thickbowel	thickened bowel in x-ray	X	X	
ascites	ascites or peritoneal fluid in x-ray	X	X	
freeipair	free intraperitoneal air	X	X	
venton	ventilation on dx			X
ventpri	No of days on ventilation prior to dx			
ventever	Ever on ventilation if no ventilation on dx?			
cpapon	CPAP on dx			
cpappri	No of days on CPAP prior to dx			
cpapever	Ever on CPAP if no CPAP on dx?			
vasspri	ever on vassopressor prior to dx?			X
vasson	vassopressor on dx			X

antion	antiobiotics on dx?			
cranult	cranial ultrasound done for ivh?			X
entnutrec	enteral nutrient on dx?			
culturefive	any pos culture within 5 days of dx			X
cultsix	any pos culture 6-14 days before dx			
wbccell	wbc		X	
neutcount	neutrophil count			
neutperc	neutroperc			
bandscount	bands count			
bandsperc	bands percentage			
platcount	platelet counts		X	
hemocrit	hematocrit			
reacpro	CRP			
phval	pH values		X	X
phsite	site where blood was collected for pH			

[00112] Three cohort sets of patient data were analyzed: (1) clinical findings on 69 patients including Bell's NEC staging criteria (Table 5); (2) urine peptidomes on 34 individual patients (Table 6); and (3) plasma proteomes on 90 individual patients (Table 7).

[00113] **Table 5:** Demographics between Non-Progressive vs Progressive NEC patients in the Clinical Assays. Mann Whitney test for continuous variables and Fischer Exact test for dichotomous variables. [ ] represents 95% confidence interval. () represents percentages.

	Non-Progressive n=43 (62.3%)	**Progressive** n=26 (37.7%)	**p-value**
Male	19 (44.2%)	18 (69.2%)	0.051
Gestational Age (week)	29.8 [28.0-30.9]	28.4 [27.0-28.9]	0.122

Birth Weight (gm)	1343.4 [1130.9-1556.0]	1164.3 [932.1-1396.6]	0.220
Race			0.315
White	21 (48.8%)	9 (34.6%)	
Black	15 (34.9%)	11 (42.3%)	
Asian	3 (7.0%)	0 (0%)	
Native Hawaiian/Pacific Islander	0 (0%)	1 (3.9%)	
American Indian/Alaskan Native	0 (0%)	0 (0%)	
Latino or Hispanic	9 (20.3%)	6 (23.1%)	1.000

[00114] **Table 6:** Demographics Among Non-Progressive and Progressive NEC patients in the Urine Assays. General Linear Model & ANOVA with Scheffe adjustment for continuous variables and Fischer Exact test for dichotomous variables. [] represents 95% confidence interval. () represents percentage.

	Non-Progressive n=17 (50.0%)	**Progressive** n=11 (32.4%)	**p-value**
Male	7 (41.2%)	10 (90.9%)	0.025
Gestational Age (week)	28.9 [27.3-30.6]	28.0 [25.9-30.1]	0.236
Birth Weight (gm)	1230.5 [917.3-1543.7]	1167.9 [778.6-1557.3]	0.609
Race			0.138
White	12 (70.6%)	4 (36.3%)	
Black	3 (17.7%)	5 (45.5%)	
Asian	2 (11.7%)	0 (0%)	
Native Hawaiian/Pacific Islander	0 (0%)	0 (0%)	
American Indian/Alaskan Native	0 (0%)	0 (0%)	
Latino or Hispanic	2 (11.8%)	3 (27.3%)	0.449

[00115] **Table 7:** Demographics between Non-Progressive vs Progressive NEC patients in the Plasma Assay. Mann Whitney test for continuous variables and Fischer Exact test for dichotomous variables. [ ] represents standard deviations. ( ) represents percentage.

	Non-Progressive n=60 (66.7%)	Progressive n=30 (33.7%)	p-value
Male	25 (43.1%)	20 (69.0%)	0.026
Gestational Age (week)	30.2 [29.1-31.3]	28.2 [26.8-29.5]	0.031
Birth Weight (gm)	1453.8 [1245.0-1662.7]	1128.4 [916.2-1340.6]	0.054
Race			0.140
White	32 (55.2%)	11 (37.9%)	
Black	17 (29.3%)	12 (41.4%)	
Asian	4 (6.9%)	0 (0%)	
Native Hawaiian/Pacific Islander	0 (0%)	1 (3.5%)	
American Indian/Alaskan Native	0 (0%)	0 (0%)	
Latino or Hispanic	11 (19.0%)	6 (20.7%)	1.000

[00116] Patient demographics analysis. Once enrolled, epidemiologic data were abstracted from the patient's chart as previously described (3) until one of several end-points was reached. Proportion and its confidence interval were employed to identify possible outliers in the non-progressive and progressive NEC patients. Fisher's exact test, Student T test and Mann Whitney U test were performed to examine the distribution of each demographic variable between non-progressive and progressive NEC patients. A general linear model with ANOVA was conducted to compare each demographic variable among the non-progressive and progressive. Scheffe adjustment was added to correct the p-values for multiple pair-wise comparisons. All analyses on the demographic variables were executed using SAS statistical software version 9.1 .3.

[00117] Urine collection, storage and processing. Intra day urine samples (0.5 ml_ ~ 1 ml_) were collected in sterile tubes and held at 4°C for up to 8 h before centrifugation (2,000 g x 20 min at room temperature) and freezing of the supernatant at -70^. The details of urine processing, preparation of peptides, extraction and fractionation are reported elsewhere (13)

[00118] Urine peptidomic MS data analysis. The ABI 4700 oracle database MS spectra were exported as raw data points via ABI 4700 Explorer software ver 2.0 for subsequent data analyses. The m/z ranges were from 800 to 4000 with peak density of maximum 30 peaks per 200 Da, minimal S/N ratio of 5, minimal area of 10, minimal intensity of 150, and 200 maximum peaks per spot. An informatics platform was previously developed which contains an

integrated set of algorithms, statistical methods, and computer applications, to allow for MS data processing and statistical analysis of liquid chromatography-mass spectrometry (LCMS) based urine peptide profiling. The MS peaks are located in the raw spectra of the matrix-assisted laser desorption/ionization (MALDI) data by an algorithm that identifies sites (mass-to-charge ratio, m/z values) whose intensity is higher than the estimated average background and the ~100 surrounding sites, with peak widths ~0.5% of the corresponding m/z value. To align peaks from the set of spectra of the assayed samples, linkage hierarchical clustering was applied to the collection of all peaks from the individual spectra. The clustering, computed on a 24 node LINUX cluster, is two dimensional, using both the distance along the m/z axis and the HPLC fractionation time, with the concept that tight clusters represent the same biological peak that has been slightly shifted in different spectra. The centroid (mean position) of each cluster was then extracted to represent the "consensus" position as the peak index (bin) across all spectra.

[00119] MS/MS analysis for peptide biomarkers. The approach of ion mapping was used to obtain protein identificatyion. In ion mapping, biomarker candidate mass spectra (MS) peaks are selected on the basis of discriminant analysis and then targeted for MS/MS sequencing analysis. Extensive MALDI-TOF/TOF and LTQ Orbitrap MS/MS analyses coupled with database searches were then performed to sequence and identify these peptide biomarkers. The identity of a subset of peptides detected was determined by searching MS/MS spectra against the Swiss-Prot database (June 10, 2008) restricted to human entries (15,720 sequences) using the Mascot (version 1.9.05) search engine. Searches were restricted to 50 and 100 ppm for parent and fragment ions, respectively. No enzyme restriction was selected. Since we were focusing on the naturally occurring peptides, hits were considered significant when they were above the statistical significant threshold (as returned by Mascot). Selected MS/MS spectra were also searched by SEQUEST (BioWorksTM rev.3.3.1 SP1) against the International Protein Index (IPI) human database version 3.5.7 restricted to human entries (76,541 sequences). mMASS, an open source mass spectrometry tool (http://mmass.biographics.cz/), was used for manual review of the protein identification and MS/MS ion pattern analysis for additional validation. Different fragmentation techniques were used for the validation of a peptide sequence, as well as for the detection, localization and characterization of post-translational modifications.

[00120] Pathway analysis. The PANTHER (Protein ANalysis THrough Evolutionary Relationships) Classification System (20) is a unique resource that classifies proteins by their functions and molecular pathways, using published scientific experimental evidence and evolutionary relationships. The protein IDs of the protein precursors of the urine peptide biomarker candidates were uploaded to PANTHER 7.0 (http://www.pantherdb.org/) to explore the molecular pathways these biomarkers might involve.

[00121] **SELDI-TOF MS, analysis and feature extraction.** Aliquots of plasma were thawed, denatured, and fractionated on an anion exchange column using the Expression Difference Mapping kit from Ciphergen Biosystems in conjunction with Beckman Biomek 2000 robot. Each plasma sample was processed in duplicate; including controls. For fractionation, 20 mL of each plasma was denatured with 30 mL 9 M urea, diluted to 1 M urea at pH 9, and applied to Q ceramic HyperD F ion-exchange beads (strong anion exchanger). The pass-through and a pH 9 wash were combined as fraction 1 by filtration of the beads in a 96-well vacuum filtration plate (Millipore, Bedford, MA, USA). Fractions 2 (pH 7), 3 (pH 5), 4 (pH 4), 5 (pH 3), and 6 (organic elution) were similarly collected. All fractions had a total volume of 200 mL per sample and were stored at -80ºC until further processing. For SELDI analysis, fractions were thawed and 10 mL aliquots of each sample were diluted ten-fold in binding buffer appropriate for the CM10 (weak cation exchanger, 0.1 M acetate pH 4.0) and H50 (RP, water:ACN:TFA 90:10:0.1), Ciphergen SELDI surfaces. Each surface was prepared according to the manufacturer's instructions and then incubated with each appropriately diluted sample. After incubation, each surface was washed successively with binding buffer and water. After brief air-drying, 1 mL of saturated sinapinic acid was added twice to each spot. Mass spectra of spotted samples were obtained using Ciphergen PBSIIc mass spectrometer. The detector voltage was set to 2900 V, laser intensity 170, and detector sensitivity. Data collection was optimized for m/z 3000-30 000, and the digitizer frequency was 250 MHz. Spectra were collected by Ciphergen ProteinChip software 3.2 and exported to CDM and feature extraction was performed using the software "Simultaneous Spectrum Analysis" (SSA) .

[00122] **Statistical analyses.** Hypothesis testing used Student t test and Mann-Whitney U test, and global and local FDR to correct for multiple hypothesis testing issues. Nearest shrunken centroid (NSC) based feature selection, including permutation based FDR analysis, was performed using R PAM package. Unsupervised heatmap analyses were performed using R stats package. Binary class clustering results were grouped into modified 2 X 2 contingency tables, which were used to calculate the proportion of the clustering results that agreed with clinical diagnosis and the statistical significance by the Fisher's exact test. Supervised linear discriminant analysis for binary (NEC M and S) classifications, using R MASS package, led to the predictive linear discriminant analysis models. The predictive performance of each linear discriminant analysis model was evaluated by ROC curve analysis. The class prediction results were grouped in modified 2 X 2 contingency tables and the statistical significance of the extent of agreement with clinical diagnosis was assessed by Fisher's exact test.

[00123] Predictive probabilities from the linear discriminant model (LDA) of NEC clinical parameter panel (11 clinical parameters) were transformed into scores.

[00124] NEC clinical score:

[00125] *X = scale (log (Clinical model LDA P value X 100)) X 10*

[00126] Scale is a generic R function whose default method centers and/or scales the columns of a numeric matrix. The scoring metrics enable the clinical parameter based classifier to be interpreted on a scale, rather than a strict binary discrimination. This increases the flexibility and the collective use of each of the panel components. NEC subjects were sorted by the corresponding NEC clinical scores (from smallest to largest) and stratified. For each patient, the percentage of NEC subjects with equal or lower score was plotted against this subject's clinical score. Visual inspection of the NEC score percentile versus the NEC clinical score plot separated the patients into low, intermediate and high-risk groups. Each group's risk of NEC progression was quantified as the proportion of NEC S class diagnoses among the group's patients.

RESULTS

[00127] **Patient demographics and characteristics.** In this study, a systematic approach was taken to discover biomarkers of NEC progression by examining three cohort sets of patient data: (1) clinical findings on 69 patients including Bell's NEC staging criteria; (2) urine peptidomes on 34 individual patients; (3) plasma proteomes on 90 individual patients. Among these different data sets, 24 patients (NEC 13 M and 11 S) had complete data for clinical findings and molecular profiles for both urine peptidome and plasma proteomes. Each cohort's sample demographics are described in Table 4, 5, and 6 of the methods section, respectively. Statistically significant differences (P value < 0.01) in patient demographics were found for gestational age and gender, each of which has been cited previously.

[00128] **Bell's NEC staging system cannot be used for NEC progression risk prediction.** The NEC staging system according to Bell (Bell's Criteria) is commonly used to diagnose and more generally characterize the severity of NEC (16). Utilizing the clinical parameters that comprise Bell's modified criteria (Bell's modified criteria: Feeding intolerance, Apneic / bradycardic episode, Oxygen desaturation episoe, Grossy bloody stools, Abdominal distention, Abdominal tenderness, Pneumatosis intestinalis, Portal venous gas, Lleus, Dilated bowel, Pneumoperitoneum, Air / Fluid levels, Thickened bowel, Ascites or peritoneal fluid, Free intraperitoneal air; clinical parameters detailed in Table 4 of the methods section), linear discriminant analysis was performed on a training data set from NEC M (n=30) and S (n=17) samples. The resultant LDA model was then tested on a new data set comprised of NEC M (n=13) and S (n=9) samples. The predicted probabilities for the progression of NEC for both the training (left) and testing data (right) were plotted (Figure 1A) for each of the patients. In Figure 1B and 1C (Figure 1B training and Figure 1C testing), samples are partitioned by the true class (upper) and predicted class (lower). The 2X2 contingency tables summarize the classification results for NEC progression. In the training set, an overall 80.9% agreement with the clinical outcome is realized using the LDA model (29/30 NEC-M and 9/17 NEC-S; P value of 1.4X10-4), however only 52.9% of the NEC S subjects were classified correctly. Using the

LDA model to analyze a new independent dataset for testing yields a mere 11.1% (1 of 9) correctly classified as progressive NEC, and an overall 63.6% agreement (p=0.41) with the clinical diagnosis (Figure 1D) when both medical and surgical outcomes are considered together. Unsupervised clustering of all 69 samples revealed no obvious pattern, supporting the findings from the supervised learning that Bell's NEC staging criteria are inadequate for predicting the risk of NEC progression.

[00129] 11 **clinical-parameter based classifier was developed for NEC patient stratifications.** Detailed clinical data for 50 distinct clinical parameters (Table 3) were collected and Mann Whitney U test was used to analyze NEC M (n=43) and S (n=26) patient groups. Eleven clinical parameters (pH value of blood), Portal venous gas in x-ray, Abdominal ileus in x-ray, use of vassopressor prior to diagnosis, Abdominal distention, Cranial ultrasound done for ivh, Vassopressor on diagnosis, Ventilation on diagnosis, Any positive culture within 5 days of diagnosis, Gestational age, Birth weight; Mann Whitney U test P value < 0.1) were selected for subsequent LDA modeling, and the corresponding absolute values (ABS) of the first linear discriminant (LD1) from the LDA were plotted. The clinical parameters of pH, portal venous gas on x-ray, abdominal ileus by x-ray, use of vasopressor medications prior to diagnosis, and abdominal distention were found to be the most distinguishing clinical parameters between NEC classifications for M and S subjects. The use of the 11 clinical parameter panel on a training (NEC 30 M and 17 S) and test set (NEC 13 M and 9 S) revealed good separation between the highest and next highest probability for the classification (Figure 2A). Overall, 28 of the 30 NEC M and 11 of the 17 NEC S in the training set, and 13 of the 13 NEC M and 6 of the 9 NEC S in the testing set were classified correctly. Overall, the 11-clinical-parameter panel classified the training and test sets with a performance P value of $2.1X10^{-4}$ (AUC of ROC: 0.927) and $1.1X10^{-3}$ (AUC of ROC: 0.923) respectively. However, the NEC S prediction rates were sub-optimal with only 64.7% and 66.6% agreeable with the clinical diagnosis.

[00130] **Urine 36-peptide panel effectively classified NEC M and S patients.** MALDI-TOF mass spectrometry (MS) based urine peptidomic analysis resulted 120 HPLC fractions for each sample, resolving a total of 17,173 peptide peaks defined by distinct m/z and HPLC fractions in the 900- to 4000-Da range. All the features were ranked by a nearest shrunken centroid (NSC) algorithm (26) in order to differentiate NEC M (n=17) and S (n=17) groups. For the NEC-S class, 6 patient samples were obtained following surgery, the reminder (n=11) were obtained at the time of diagnosis, same as the samples for the NEC-M class patients. Next, the top 1000 peaks were subject to extensive MSMS protein identification yielding 473 distinct peptides. Unsupervised cluster and pathway analyses of these identified urine peptides were performed for the NEC M (n=17), S (n=11) and post surgical (Post S, n=6) subjects. Manual examination of the heat map display of unsupervised clustering revealed that the 473 urine

peptides can be largely grouped into 2 bins: (I) peptides up regulated in NEC S, then down in NEC M; (II) upregulated in NEC M, down in NEC S samples. Data mining software (Ingenuity Systems, www.ingenuity.com, CA) was used to analyze these differential urine peptides' parent proteins and to identify significant gene ontology groups and relevant signaling pathways. As shown in Figure 3B, the analysis of significance (-log(P)) of the canonical pathways could largely group them into 3 bins: (1) similarly significant in NEC M and S: atherosclerosis, dendritic cell maturation, notch signaling; (2) more significant in NEC M than S: hepatic fibrosis/hepatic stellate cell activation, caveolar-mediated endocytosis signaling, virus entry via endocytic pathways; (3) more significant in NEC S than M: coagulation system, acute phase response signaling.

[00131] Sequence analysis of these NEC differentiating urine peptides, and their relative abundance represented by NSC values (26), revealed that several came from the same precursor proteins, and included (Figure 3C) collagens (COL1A1, COL1A2, COL2A1), epithelial-mesenchymal cell interaction (EMI) domain-containing protein 1 (EMID1), Eps 15-Homology (EH) domain-binding protein 1-like protein 1 (EHBP1 L1), fibrinogen alpha chain (FGA), gliding motility protein gliomedin (GLDN), hemoglobin subunit alpha (HBA1), Teneurin-3, PRAGMIN, steroidogenic factor 1 (SF1), and uromodulin (UMOD).

[00132] To develop a biomarker panel with manageable panel size, we built LDA classifiers with various subsets of the top ranked (NSC algorithm), therefore most significant, 473-peptide (sequence identified through MSMS analysis) data set. From these differentially expressed urine peptides, we sought to identify a biomarker panel of optimal feature number, balancing the need for small panel size, accuracy of classification, goodness of class separation (NEC M vs S), and with sufficient sensitivity and specificity. Goodness of separation is defined by computing the difference (Ä) between discriminative scores, calculated as estimated probabilities (13, 26). When class is predicted correctly, Ä probability is the difference of the highest and next highest probability; when predicted incorrectly, Ä probability is the difference of the probability of the true class and the highest probability, which will be negative. In Figure 4A are the NEC M and S box-whisker graphs. Boxes contain 50% of values falling between the 25th and 75th percentiles; the horizontal line within the box represents the median value and the "whisker" lines extend to the highest and lowest values. This analysis revealed 36 peptides to be the smallest panel size for which the "box" values of goodness of separation are positive for both NEC M and S. To assess the association of the disease status with the abundance pattern of these 36 peptides, we performed unsupervised hierarchical cluster analysis with heat map plotting (Figure 4B). The analysis demonstrated 2 major clusters reflecting NEC disease progression status, reinforcing the effectiveness of this 36-urine-peptide "signature" in predicting NEC M and S class distinction. Student T test and Mann-Whitney U test, in addition to MSMS sequence identification analyses (Figure 4C) were performed for these 36 urine

peptides. Close examination of these 36 peptides revealed nested peptides for COL1A2 (m/z 1853, 1752), COL1 1A2 (m/z 1529, 1679), FGA (m/z 1568, 2560, 2659), and UMOD (m/z 1680, 1912) having overlapping sequences derived from the same parent protein precursors. Further pathway analysis (Figure 4D) using the PANTHER database (20) revealed these 36 peptide biomarkers derived from protein precursors involved in integrin signaling pathway (65.7%, P00034), plasminogen activating cascade (11.4%, P00050), blood coagulation (11.4%, P0001 1), ubiquitin proteosome pathway (8.6%, P0060), and inflammation mediated by chemokine and cytokine signaling pathway (2.9%, P00031) respectively. These findings are consistent with the presumed pathophysiology of exuberant inflammatory reaction resulting in coagulative necrosis of the gut wall.

[00133] Plasma Protein Panel yields effective class prediction for NEC M and S patients. Patient blood samples were subject to SELDI-TOF MS based plasma proteomic analysis (21) that resolved a total of 1528 protein peaks. All protein peaks were ranked by a nearest shrunken centroid (NSC) algorithm differentiating NEC M (n=60) and S (n=30) groups. As above, we sought to identify a biomarker panel of optimal features to achieve goodness of class separation (NEC M vs S), and with sufficient sensitivity and specificity. We built LDA classifiers with various subsets of the 1528-protein-peak data set. The computed goodness of separation (Figure 5A) (defined above) is shown in Figure 5A as the NEC M and S box-whisker graphs. As before, the boxes contain the interquartile range of values, the horizontal line within the box represents the median value and the "whiskers" extend to the highest and lowest values. This analysis revealed 48 to be the smallest panel size for which the "box" values of goodness of separation are positive for both NEC M and S. A close examination of the spectra revealed that these 48 spectral peaks actually are from 30 unique proteins (Figure 5B). Relative abundance of the 30 plasma proteins (Figure 5B) were analyzed by the nearest shrunken centroid values in either NEC M or S patient class with Color Scale conditional formatting. The significance of each plasma protein biomarker was quantified by Mann-Whitney U test and Student T test P values, demonstrating (reflecting) each plasma protein's individual effectiveness as a biomarker in differentiating NEC M from S groups. To assess the association of the disease status with abundance patterns of these 30 plasma proteins, we performed an unsupervised hierarchical cluster analysis with heat map plotting (Figure 5C). The analysis shows NEC subjects clustered largely according to the disease progression status, reinforcing the effectiveness of this plasma-protein-peak "signature" in differentiating NEC M and S.

[00134] Comparative analyses of clinical and molecular (urine/plasma) based biomarker panels via unsupervised learning. To compare the discriminant performance of different biomarker panels comprised of either 11 clinical parameters, 36 urine peptides, or 30 plasma proteins, a set of NEC patients (13 M and 11 S) were selected for which complete datasets of

clinical findings, and urine/plasma profiling were available. Unsupervised cluster analyses were applied to determine how the NEC subjects were organized according to these clinical or molecular based biomarker classifiers. As shown in Figure 6, each biomarker panel's differentiating pattern was represented by a corresponding cluster heat map. Recognizing the branch with the largest number of the clustered NEC S subjects as the NEC S "class" and the remaining as the NEC M "class", the unsupervised discriminating significance of these different biomarker panels was quantified by the Fisher exact test of the 2X2 tables partitioning the clinically known subjects by the cluster grouping: clinical parameter panel (11 features), P value 0.64; urine peptide panel (36 features), P value 9.5X10-4; and plasma protein based panel (30 features), P value 0.01. The 36-urine-peptide panel appeared to be more effective than the 11-clinical-parameter or the 30-plasma-protein panel in discriminating NEC M from S subjects.

[00135] Integrative analyses of clinical and molecular (urine/plasma) findings reveals an optimal biomarker panel of 15 urine peptides and 3 plasma proteins for NEC progression. Through the unsupervised analysis, we were exploring whether an analysis integrating the clinical, urine peptide and plasma protein based biomarkers can achieve better predictive accuracy in NEC progression analysis. As shown in Figure 6D, overall, with a P value of $5.2X10^{-4}$, the combined panel of 11 clinical parameters, 36 urine peptides and 30 plasma proteins correctly clustered 92.3% of NEC M and 81.8% NEC S subjects respectively, indicating greater effectiveness in NEC progression prediction for the integrated approach over any of the individual classifiers.

[00136] To find a predictive biomarker panel of optimal and manageable feature number, various subsets out of the combined biomarkers from different sources were tested as classifiers to analyze both their goodness of separation and false discovery rate (FDR). Linear discriminant probabilities of a biomarker panel of 18 features were found to be optimal for goodness of separation of the NEC M and S subjects (Figure 7A). The FDRs of the LDA classifiers were estimated and were shown to significantly increase after the feature size expanded to greater then 18 (Figure 7B). Therefore, the 18-feature biomarker panel was chosen as the optimal biomarker set, balancing the need for small panel size, accuracy of classification, goodness of class separation (NEC M versus S), and sufficient sensitivity and specificity. This 18-biomarker panel consisted of 15 urine peptides (corresponding to 13 proteins Q6ZUQ4, OBFC2B, COL11A2, NBEAL2, GRASP, HUWE1, COL1A2, HOXD3, DSG4, KRTAP5-11, Y1020, FGA, UMOD; close examination of the 15 urine peptides again revealed the overlapping peptide fragments of FGA (MW: 2559, 2659) and UMOD (MW: 1679, 1911)) and 3 plasma peptides (CTAPIII, SAA1, B2M, TTR) The relative abundance of the 18 peptide biomarker panel (Figure 7C) was analyzed by the nearest shrunken centroid values in either NEC M or S patient class and plotted with Color Scale conditional formatting.

[00137] An unsupervised analysis by heat map plotting across the 18 biomarkers demonstrated that all 11 of 13 NEC M subjects and importantly 10 of 11 NEC S subjects clustered together, co-clustered (Figure 7D). The overall clustering agreement with clinical diagnosis is 87.5% and discriminant significance (P value) is 6.4 X 10^{-4}. Using the18-biomarker data set, supervised analysis was performed to develop the LDA model and the estimated probabilities were plotted (Figure 7E). Samples were partitioned by the true class (upper) and predicted class (lower). The 2X2 contingency tables (Figure 7E) summarizes the NEC M/S classification results, which are 100% agreeable with clinical diagnosis and P value of 4.0X10-7 by Fisher exact test. However, in order to avoid the problem of overfitting and bias, a bootstrapping method was used to resample the original 18-biomarker data set (NEC 13 M and 11 S subjects) 500 times, thus creating 500 new sets for LDA modeling and subsequent testing. For each of the bootstrapping sets, we used the LDA derived prediction scores of each sample to construct ROC curves. To summarize the 500 ROC analyses (Figure 7F), box and whisker plots were used to describe the vertical spread around the median, and then the vertical average of the 500 ROC curves was plotted (dashed line). The ROC analyses yielded an average AUC of 0.99, demonstrating the robustness of the 18-biomarker panel in the discrimination of the NEC M and S class subjects.

[00138] **A sequential ensemble analysis of the clinical and molecular biomarker classifiers for practical and effective prediction of NEC progression.** Ensemble Data Mining Methods, also known as Committee Methods or Model Combiners, were used to combine the clinical and molecular biomarker classifiers in order to derive practical algorithms for NEC management. These machine learning methods leverage the power of multiple models to achieve better prediction accuracy than is possible with any of the individual models on their own. We integrated the molecular classifiers, either the 36 urine based or the final 18 (15 urine peptides and 3 plasma proteins) biomarker panel, with readily available clinical data. Using the 24 NEC subjects (13 M and 11 S) of which complete datasets are available, a simulation scenario - "NEC simulation set" was undertaken.

[00139] Based upon the multivariate analysis of the 11 clinical parameters of NEC 43 M and 26 S subjects (Figure 2), NEC clinical scores were calculated ranging from -10 to 50 with a higher score indicating a greater chance or risk of NEC S. As shown in Figure 8A, each particular sample's risk of being classified as NEC S was quantified by the proportion of NEC S samples with score less than that sample's clinical score in all NEC S samples. Therefore, all NEC samples were divided into low, intermediate, and high-risk groups based on their scores. A NEC clinical score of less than 20 classified samples into the low risk group, which produced a perfect match for the sub-group diagnosed as NEC M subjects (26 infants). A score of 42 or greater identified the high-risk group, in which all 16 infants were diagnosed as NEC S subjects. The remaining samples were grouped into the intermediate risk group, in which 17

were NEC M and 10 were NEC S subjects. Within the intermediate risk group, there are no clear delineations between NEC M and S subjects based simply on score. Therefore, we conclude that using the NEC clinical score, it is possible to stratify the NEC subjects into low (0%), intermediate (37%), and high-risk (100%) groups. If validated to be consistently demonstrable for NEC risk stratifications, the clinical score based forecast of the NEC subjects, particularly those into the low and high-risk bins, may be clinically useful to treat according to these prognostic indications.

[00140] Close examination of these subjects with comprehensive clinical, urine peptidomics and plasma proteomics data sets (Figure 8B) found 6 in low, 11 in intermediate, and 7 in high-risk groups. These low or high risk subjects were ultimately diagnosed either as NEC M or S, reinforcing the notion there is a parallel relationship between the clinical diagnosis and the patient stratification by the NEC clinical scoring system. When tested further with either the 36 urine peptide panel or the final 18 biomarker panel, the classifications of these subjects in either low or high risk groups were in complete agreement with the clinical diagnosis, suggesting further molecular testing may be unnecessary due to the effective patient stratification by the NEC clinical based scoring system. However, as for the subjects in the intermediate groups assigned upon the NEC clinical scores, additional tests are needed to accurately classify the subjects and to predict NEC progression. For the NEC 7 M and 4 S subjects in the intermediate risk group, either the 36-peptide urine panel or the final 18-biomarker panel, classified them correctly with 100% agreement with clinical diagnosis and with a P value of 0.003 by the Fisher exact test. The simulation data set analysis suggests that the sequential and ensemble integration of the clinical and molecular based panels can adequately stratify patients to allow effective NEC management: (1) the low and high risk patients are correctly stratified for NEC progression by the clinical score; (2) the clinically intermediate risk patients are be subject to additional molecular based testing to produce further stratification thus allowing for the sensitive and specific prediction of NEC progression.

DISCUSSION

[00141] Necrotizing Enterocolitis (NEC) is a devastating inflammatory disease that affects at risk premature newborns in an un-predictable manner. NEC is a principal source of overall premature neonate mortality as well as short and long-term morbidity in surviving infants (7, 28). In general, NEC occurs in two forms that can be loosely described as non-progressive and progressive. These descriptive terms reflect the underlying degree of tissue injury that includes irreversible intestinal necrosis requiring its surgical removal. Despite numerous previous efforts, clinical parameters (3) and serologic tests alone (9, 11) appear to be inadequate for either diagnosing or predicting the outcome of NEC until late in the course of disease. Moreover, clinical signs of NEC, e.g. the x-ray finding of air or gas in the gut wall (pneumatosis

intestinalis), are both non-specific of disease progression and vulnerable to observer variability and subjective assessment. Thus, the current approach to decision-making in treating NEC is generalized, non-specific and highly observer dependent. This is problematic, since 50% of cases will remain limited, and resolve with supportive care, while an additional 30-50% progress and require surgery. This leads to a number of both under and over-treated infants with likely effects on overall outcome (29). Novel therapeutic strategies that may ameliorate or halt progression of the disease cannot currently be tested since the only reliable signs of progressive NEC occur late in the course of disease when tissue destruction is irreversible and as such meaningful changes in patient care would therefore be unlikely of increased benefit. Moreover, since not all institutions caring for infants with NEC or at risk for NEC can offer surgery as a treatment (only highly specialized centers with neonatal and pediatric surgical sub-specialists), if those infants that are most likely to progress could be identified earlier, an option for transfer to a higher level of care center would be highly advantageous, and conversely, many transfers of infants not requiring surgery could be averted. In this study we sought to address these challenges and have combined the novel use of available clinical data to effect an initial risk stratification of infants with NEC along with protein biomarker discovery. We report that the subsequent combination of these disparate datasets provides a useful and meaningful algorithm that correctly predicts NEC progression prior to the time at which obvious clinical signs of advanced disease are present. We conclude that this type of integrated and ensemble algorithm may overcome similar challenges encountered in other rare diseases that evolve either spontaneously or in response to therapy.

[00142] Like many other human diseases, NEC affects an organ system that is not readily amenable to biopsy to arrive at a definitive tissue diagnosis or prognosis. Thus, similar to other diseases, surrogate markers of disease (e.g. x-ray findings of pneumatosis intestinalis) or systemic signs (acidosis, white blood cell count) are currently utilized to risk stratify patients clinically. Various mass spectrometry based proteomics platforms are being increasingly applied to analyze available specimens (blood, urine, stool) in order to identify molecular markers of disease (biomarkers). In the current study, a robust set of several urine peptide biomarkers and plasma protein biomarkers enabled the accurate discrimination between NEC M and S urine samples. Several of these peptides were found to be derived from the same parent protein. The finding of nested peptides is both reassuring and potentially informative since it would be unusual to discover various cleavage forms from the same parent protein as a spurious finding. Moreover, the nested peptides also suggest some novel aspects of the underlying biology of NEC. For example, since several of the identified peptides are derived from various collagens, collagen 1A2 (COL1A2), collagen 11A2 (COL1 1A2), this may reflect the possible involvement of specific exo-and endo-peptidases acting on the extra-cellular matrix (ECM) and potentially contributing to the underlying pathophysiology. Also interesting is

the finding of COL4A2 (basement membranes), MUC15, and MUC3A (cell surface glycoproteins expressed in enterocytes) all with increased relative expression in the NEC S class of patients. Together, these peptides more specifically point toward a destructive process in the gut with perhaps cell surface or basement membrane breaching of the intestinal epithelium which has been proposed by several authors as contributing significantly to the pathogenesis of NEC (6, 31). One persistent finding that consistently survived all of the analyses was that of increased FGA (fibrinogen, alpha chain) peptides in the NEC S class patient urine. FGA is involved in tissue injury and blood coagulation as the most abundant component of thrombus formation. Liquefaction necrosis with significant small vessel thrombosis is a common pathologic finding in surgical NEC. In addition, various cleavage products of fibrinogen can regulate cell adhesion, display vasoconstrictor and chemotactic activities, and are mitogens for several cell types. Other significant collagens of potential biologic significance include collagen 8A1 (COL8A1) a component of vascular endothelium, and collagen 18A1 (COL18A1), also involved in the coagulation cascade. The consistent finding of peptides derived from uromodulin, the most abundant protein in normal urine, suggests a systemic inflammatory injury since uromodulin is not derived from the plasma, but rather is produced in the glomeruli. The proteolytic cleavage of an ectodomain of uromodulin on the luminal surface in the loop of Henle and its urinary secretion suggests secondary systemic effects as a result of the remote gut disease. Together, these various peptides suggest that peptide biomarkers may serve as surrogates of disease-related protease/protease inhibitors (e.g. TIMP1, MMPs) that may be differentially active in the two classes of NEC thereby reflecting the underlying tissue destruction. For example, the identification of urine peptide biomarkers suggests that active degradation of collagen is associated with the pathophysiology of NEC progression. This is in line with our previous findings that nested urinary peptide biomarkers may be generated by disease-specific exo-peptidase activity (32).

[00143] We recognize several limitations to the current study. First, a relatively small sample size was used for the integrative analyses since complete records of both molecular and clinical findings were only available for this subset of patients in the database. Future prospective studies with larger sample sizes will be needed to validate our proposed predictive algorithm. Second, the rapid identification of the urine peptide fragments using technology available in a hospital laboratory will need to be devised. Alternatively, a smaller biomarker panel and immune based detection methods will need to be developed for the reproducible detection of biomarkers following validation studies. Third, biochemical purification efforts are needed to identify those SELDI derived plasma biomarkers. Nevertheless, our application of a method integrating urine peptidomic, plasma proteomic and clinical findings for NEC risk prediction has important implications for additional disease prognosis. This approach is

especially attractive for the diagnosis and prognosis of pediatric patients in whom blood may be difficult to obtain, or is available in limited quantities.

[00144] The present sequential ensemble analyses leverages the power of the findings of both the molecular (urine peptidome and plasma proteome) and the clinical parameters based biomarker panels to achieve better accuracy in predicting the progression of NEC to an advanced stage of disease. The derivation of the scoring metrics for NEC clinical parameter-based predictions further enable the biomarker panel to be interpreted on a scale, which increases the flexibility of the panel to quantify the risk of NEC progression. The complementary effectiveness of our integrative diagnostic analysis may reflect the complex pathophysiology of NEC with diverse and interdependent clinical and biological variables. Our analyses and algorithm also suggest a potential strategy to be utilized for numerous other diseases. Diseases that can be stratified by clinical parameters and then further sub-stratified by validated biomarkers may particularly benefit. Taken together this approach and the findings presented demonstrate the additive power of integrating data from various sources.

Example 2

[00145] Bottom-up urine proteomics discovered an eleven-protein biomarker panel that effectively discriminates NEC M from S subjects.

[00146] Within the NEC samples there are 47 NEC M and 24 S subjects, which have been analyzed by mass spectrometry (MS) based urine proteome profiling using a bottom up approach. Cross validation and false discovery (FDR) guided feature selection analysis found a eleven-protein panel (PLSL, LMAN2, OSTP/OPN, APOA4, C08G, SAP, ANGT, CD14, FIBA, PROF1 , PEDF) (Fig 11), which effectively classified the NEC S samples (PLSL and LMAN2 elevated) and M samples (OSTP/OPN, APOA4, C08G, SAP, ANGT, CD14, FIBA, PROF1 , and PEDF elevated) with overall 85.9% accuracy (P value $2.6X10^{-8}$, ROC AUC 92.3% (Fig. 12). Intriguingly, several of these proteins have known biologic functions that may be related to the pathogenesis of progressive NEC and therefore reflect the underlying biology. The PRP CD14 and immune-modulating properties of PEDF were discussed above. Additionally, osteopontin (OSTP/OPN) is a phosphoprotein with a range of described biologic functions including as a pro-inflammatory cytokine for monocytes and macrophages, as well inhibitory of macrophage nitric oxide production. Fibrinogen A (FIBA), a potent member of the coagulation cascade appears to be highly expressed in NEC S class consistent with the high level of coagulative and consumptive necrosis that occurs in advanced cases of NEC (NEC S). Most intriguingly, we also found two peptide fragments from FIBA in the 36-member classifier for progressive NEC above (not shown) thus providing further support for the involvement of this molecule in NEC progression and its utility as a biomarker of progressive disease (NEC S).

Example 3

[00147] Bottom-up urine proteomics discovered a seven-protein biomarker panel that effectively discriminates NEC from Sepsis subjects. We sought to identify protein biomarkers of NEC that exist in the urine of infants at the time of first clinical suspicion of either NEC or sepsis. An un-biased, high-throughput proteomic discovery approach was taken utilizing subject samples that were obtained by the NEC Consortium. 71 NEC and 13 Sepsis urine samples underwent mass spectrometry (MS) based urine proteome profiling using a bottom up approach. Each proteome was fragmented by trypsin digestion. Full mass spectrometry scan was acquired on an LTQ FTMS, which was followed by MS/MS analysis. Protein identification was performed by searching Swiss-Prot database. Quantification of proteins in different samples was done by means of spectral counting, implementing the recent S1N algorithm . From the MSMS protein identifications, a separate list of proteins was created for each sample, and the lists were then compared to find differentially expressed proteins. For any given protein, the significance of the relative abundance between NEC and Sepsis groups was computed by Student's T test. Urine proteins with low P values discriminating NEC and Sepsis were explored by exploratory box-whisker plot analysis. Cross validation and false discovery (FDR)guided feature selection analysis revealed a seven-protein panel (CD14, SAP1 , PEDF, ftsY, PROC, MAPI B, CSN5) (Fig 11) that effectively classified the NEC and Sepsis samples with overall 95.2% accuracy (P value $1.9X10^{-9}$, ROC AUC 93% Fig. 10). Among the identified proteins as biomarkers of NEC include several that may be of particular interest given their described biologic functions and the prevailing hypothesis of NEC etiology that includes enteric bacterial invasion of the newborn gut and the inciting inflammatory cascade that results in coagulative necrosis. Perhaps most interesting is CD14, an integral part of the innate immune system as a pattern recognition receptor (PRP) that acts as a co-receptor along with Toll-like receptor 4 (TLR4) and has been implicated as causative of NEC. Although the primary ligand of CD14 is bacterial LPS, it also recognizes other pathogen associated molecular patterns. CD14 exists in two forms including a soluble form (sCD14) that can be shed or secreted from enterocytes. In addition, PEDF (pigment epithelial derived factor) is a serine protease glycoprotein that is known to effect macrophage function through PPAR* and may therefore play a role in modulating NEC associated inflammation.

[00148] The preceding merely illustrates the principles of the invention. It will be appreciated that those skilled in the art will be able to devise various arrangements which, although not explicitly described or shown herein, embody the principles of the invention and are included within its spirit and scope. Furthermore, all examples and conditional language recited herein are principally intended to aid the reader in understanding the principles of the invention and the concepts contributed by the inventors to furthering the art, and are to be construed as

being without limitation to such specifically recited examples and conditions. Moreover, all statements herein reciting principles, aspects, and embodiments of the invention as well as specific examples thereof, are intended to encompass both structural and functional equivalents thereof. Additionally, it is intended that such equivalents include both currently known equivalents and equivalents developed in the future, i.e., any elements developed that perform the same function, regardless of structure. The scope of the present invention, therefore, is not intended to be limited to the exemplary embodiments shown and described herein. Rather, the scope and spirit of the present invention is embodied by the appended claims.

CLAIMS

That which is claimed is:

1. A method of diagnosing NEC in a patient, the method comprising:

a. detecting the level in the urine of protein encoded by one or more NEC-Dx genes to obtain an NEC-Dx expression signature;

b. comparing the NEC-Dx expression signature to an NEC-Dx expression signature from a reference sample; and

c. employing the results of the comparison to provide an NEC diagnosis to the patient.

2. The method according to claim 1, wherein the one or more NEC-Dx genes is selected from the group consisting of CD14, SAP1 , PEDF, Q6ZUQ4, OBFC2B, COL1 1A2, NBEAL2, GRASP, HUWE1 , COL1 A2, HOXD3, DSG4, KRTAP5-1 1, Y1020, FGA, CTAPIII/PPBP, SAA1 , B2M, TTR, OSTP/OPN, APOA4, C08G, ANGT, FIBA, PROF1 , UMOD, PLSL, and LMAN2.

3. The method according to claim 1, further comprising obtaining an NEC clinical score, wherein the comparing comprises comparing the NEC-Dx expression signature and the NEC clinical score to an NEC-Dx expression signature and an NEC-Dx clinical score from a reference sample, and the employing comprises employing the results of both comparisons to provide a diagnosis of NEC.

4. The method according to claim 1, wherein the patient is suspected of having NEC, intestinal perforation (IP), or sepsis.

5. A method of diagnosing sepsis in a patient, the method comprising:

a. detecting the level in the urine of protein encoded by one or more sepsis-Dx genes to obtain a sepsis-Dx expression signature;

b. comparing the sepsis-Dx expression signature to a sepsis-Dx expression signature from a reference sample; and

c. employing the results of the comparison to provide a sepsis diagnosis to the patient.

6. The method according to claim 5, wherein the one or more sepsis-Dx genes selected from the group consisting of ftsy, PROC, MAPI B and CSN5.

7. The method according to claim 5, wherein the patient is suspected of having

NEC or sepsis.

8. A method of providing a prognosis for a patient with NEC or predicting responsiveness of a patient with NEC to medical therapy, the method comprising:

a. detecting the level in the urine of protein encoded by one or more NEC-M and/or NEC-S genes to obtain an NEC-M/S expression signature;

b. comparing the NEC-M/S expression signature to an expression signature from a reference sample; and

c. employing the results of the comparison to provide a prognosis for the patient or predict responsiveness of the patient to medical therapy.

9. The method according to claim 8, wherein the one or more NEC-S genes is selected from the group consisting of Q6ZUQ4, OBFC2B, COL1 1A2, NBEAL2, GRASP, HUWE1 , COL1A2, HOXD3, DSG4, KRTAP5-1 1, Y 1020, FGA, OSTP/OPN, APOA4, C08G, SAP1 , ANGT, CD14, FIBA, PROF1 , and PEDF.

10. The method according to claim 8, wherein the one or more NEC-M genes are selected from the group consisting of UMOD, PLSL, and LMAN2.

11. The method according to claim 8, wherein the patient is diagnosed as having NEC.

12. The method according to claim 8, wherein the medical therapy is antibiotics and nothing by mouth.

13. The method according to claim 8, further comprising the step of obtaining an NEC clinical score, wherein the comparing comprises comparing the NEC-M/S expression signature and the NEC clinical score to an NEC-M/S expression signature and an NEC clinical score from a reference sample, and the employing comprises employing the results of both comparisons to provide a prognosis or predict responsiveness of an NEC patient to medical therapy.

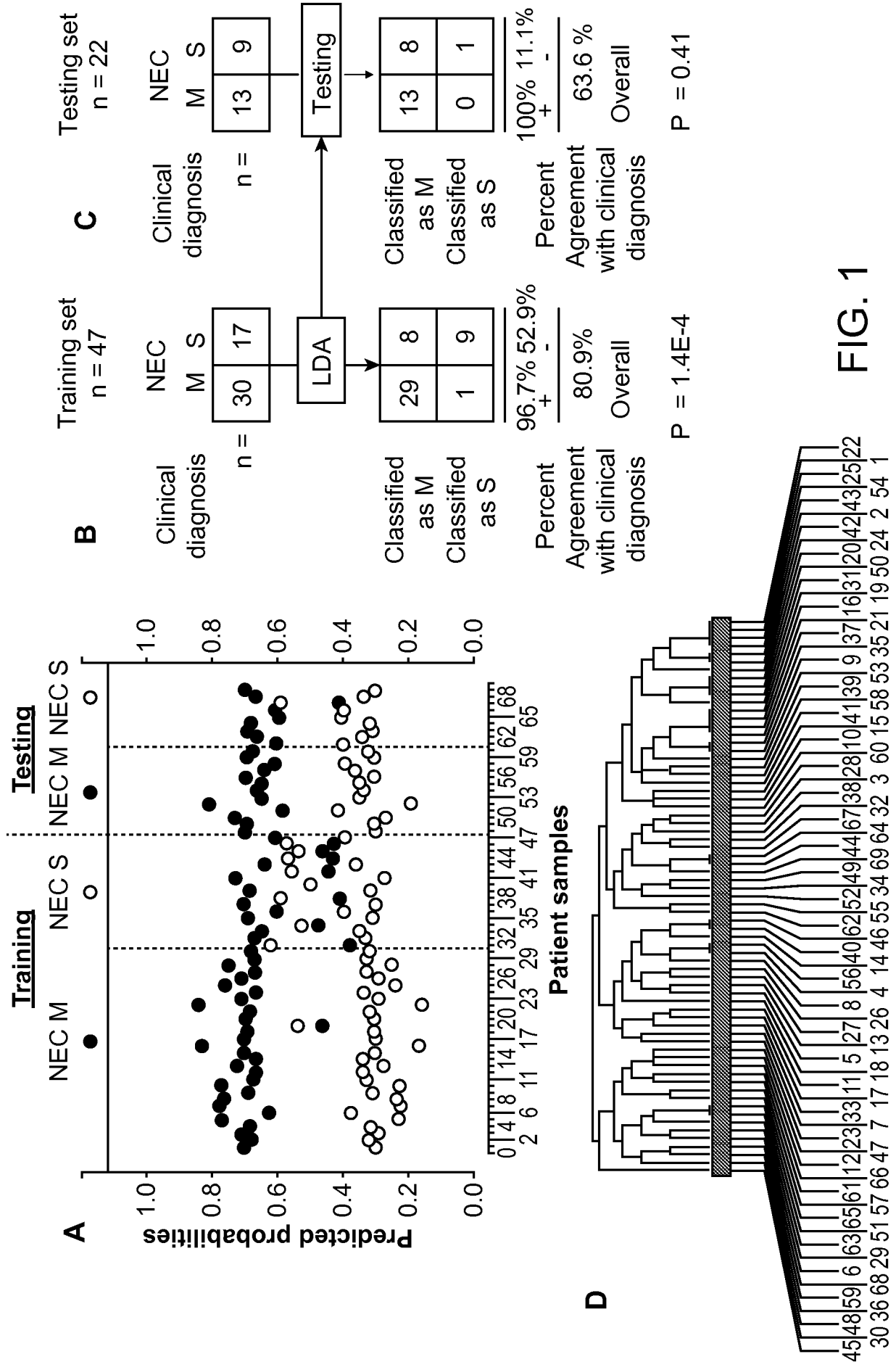
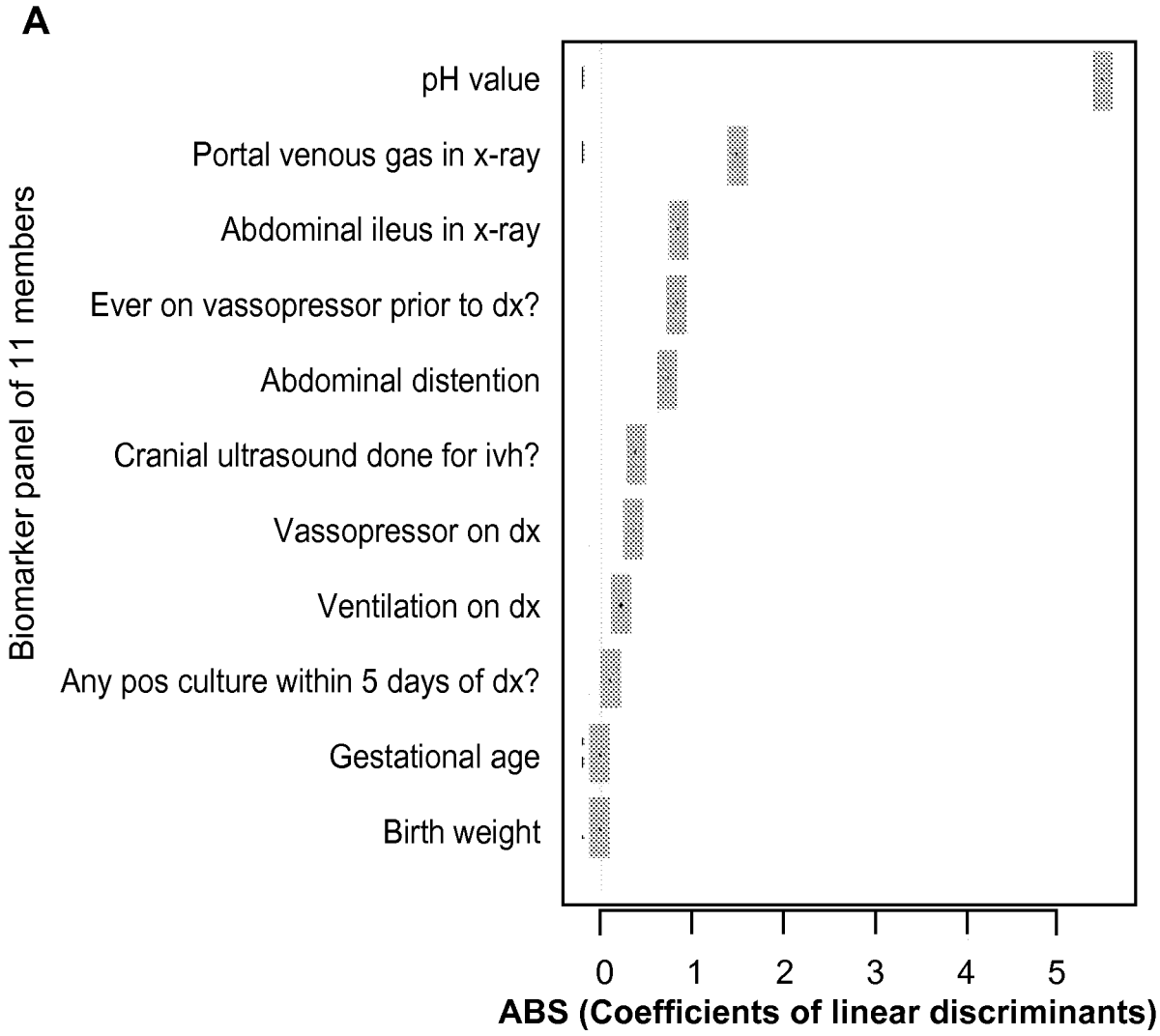

A
Predicted probabilities
Training
Testing
NEC M
NEC S
NEC M
NEC S
Patient samples
B
Training set
n = 47
Clinical diagnosis
NEC
M
S
n =
30
17
LDA
Classified as M
Classified as S
29
8
1
9
96.7%
52.9%
Percent Agreement with clinical diagnosis
+
-
80.9%
Overall
P = 1.4E-4
C
Testing set
n = 22
Clinical diagnosis
NEC
M
S
n =
13
9
Testing
Classified as M
Classified as S
13
8
0
1
100%
11.1%
Percent Agreement with clinical diagnosis
+
-
63.6 %
Overall
P = 0.41
D

FIG. 1


A
pH value
Portal venous gas in x-ray
Abdominal ileus in x-ray
Ever on vassopressor prior to dx?
Abdominal distention
Cranial ultrasound done for ivh?
Vassopressor on dx
Ventilation on dx
Any pos culture within 5 days of dx?
Gestational age
Birth weight
Biomarker panel of 11 members
0 1 2 3 4 5
ABS (Coefficients of linear discriminants)

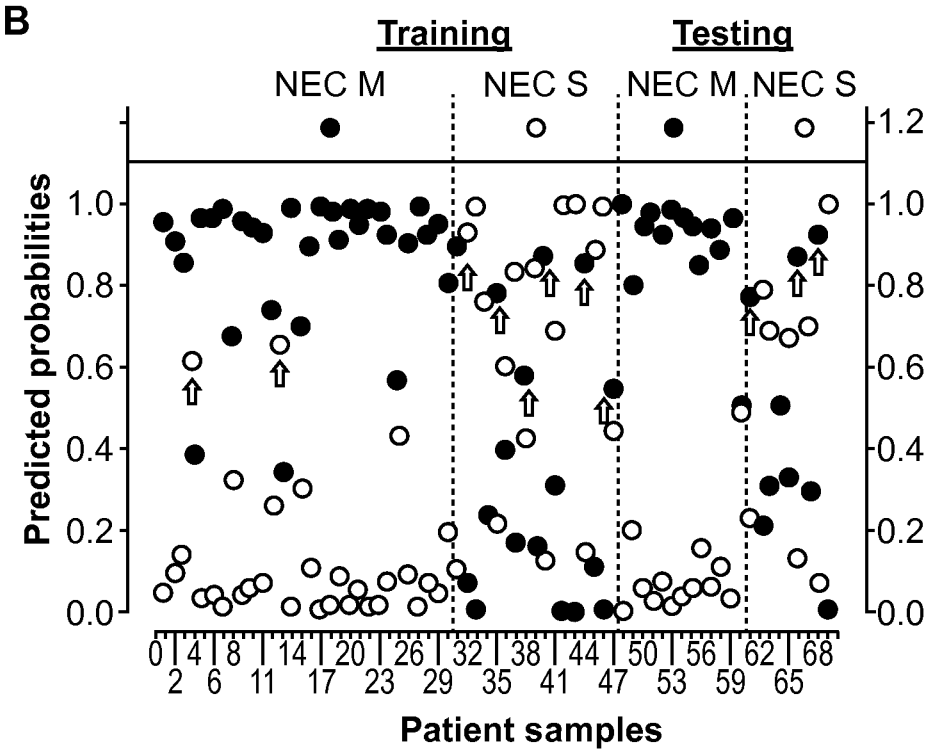
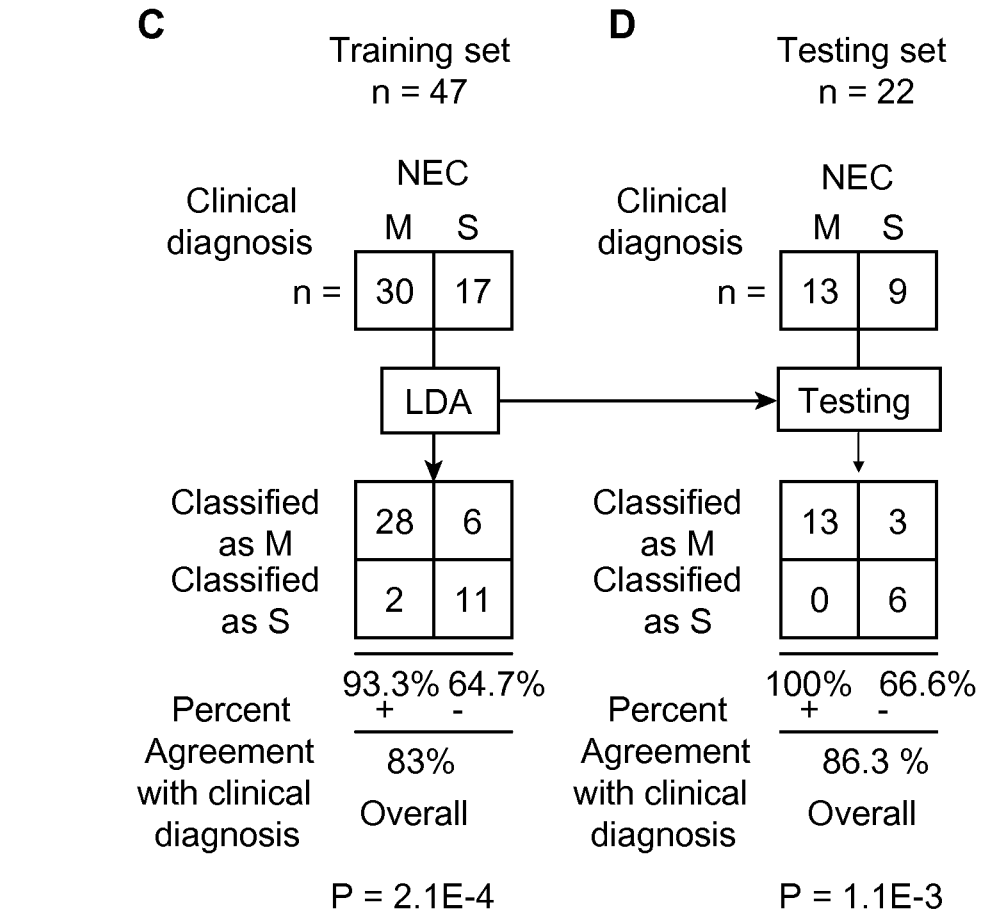

B
Training
Testing
NEC M
NEC S
NEC M
NEC S
Predicted probabilities
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0 4 8 14 20 26 32 38 44 50 56 62 68
2 6 11 17 23 29 35 41 47 53 59 65
Patient samples

FIG. 2


C
Training set
n = 47
NEC
Clinical diagnosis
M S
n = 30 17
LDA
Classified as M 28 6
Classified as S 2 11
93.3% 64.7%
+ -
Percent Agreement with clinical diagnosis
83%
Overall
P = 2.1E-4
D
Testing set
n = 22
NEC
Clinical diagnosis
M S
n = 13 9
Testing
Classified as M 13 3
Classified as S 0 6
100% 66.6%
+ -
Percent Agreement with clinical diagnosis
86.3 %
Overall
P = 1.1E-3

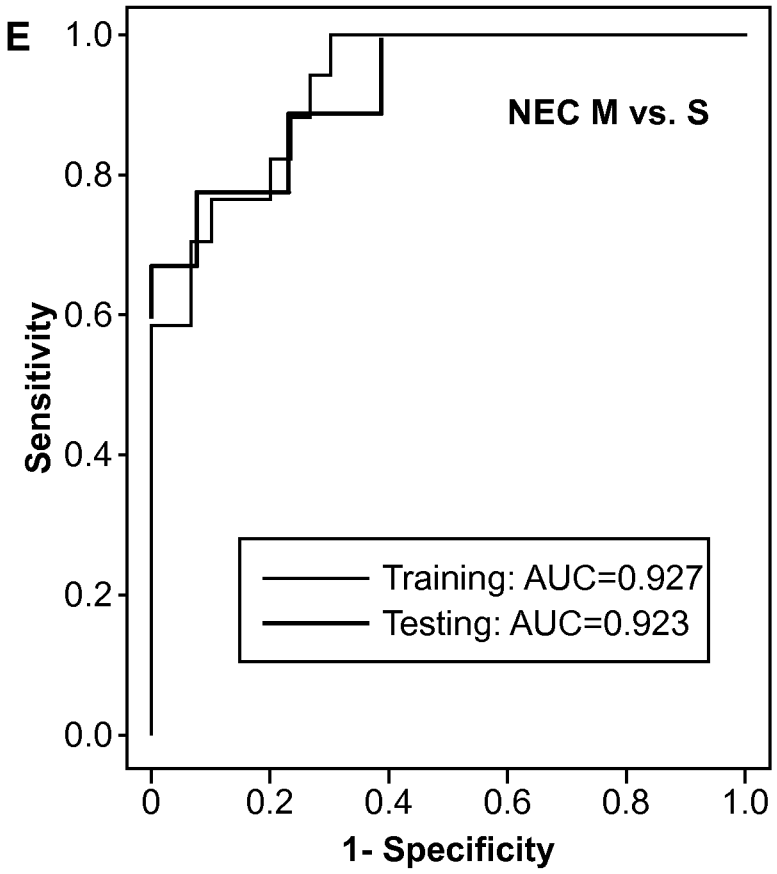

E
NEC M vs. S
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
0
0.2
0.4
0.6
0.8
1.0
1- Specificity
Training: AUC=0.927
Testing: AUC=0.923

FIG. 2 (Cont.)

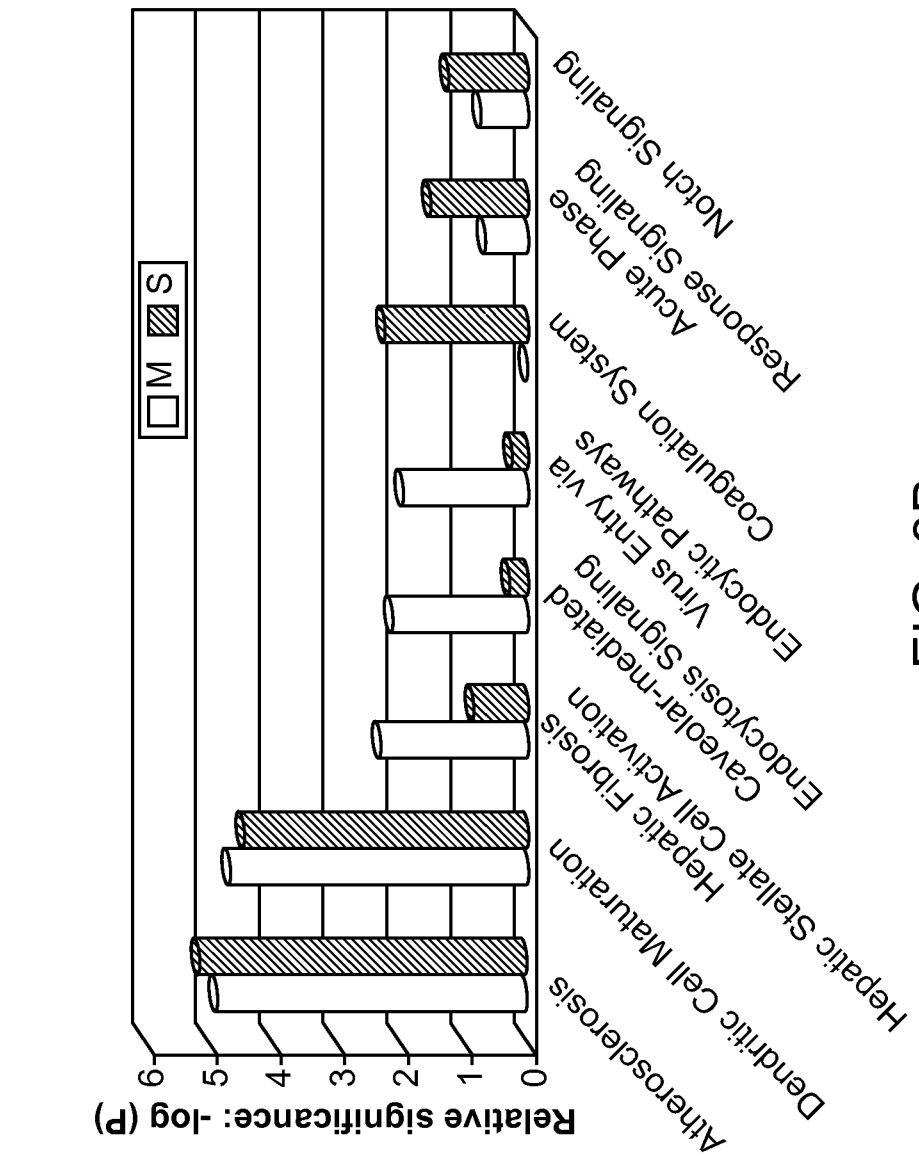

Relative significance: -log (P)
0
1
2
3
4
5
6
M
S
Atherosclerosis
Dendritic Cell Maturation
Hepatic Fibrosis
Hepatic Stellate Cell Activation
Caveolar-mediated Endocytosis Signaling
Virus Entry via Endocytic Pathways
Coagulation System
Acute Phase Response Signaling
Notch Signaling

FIG. 3B

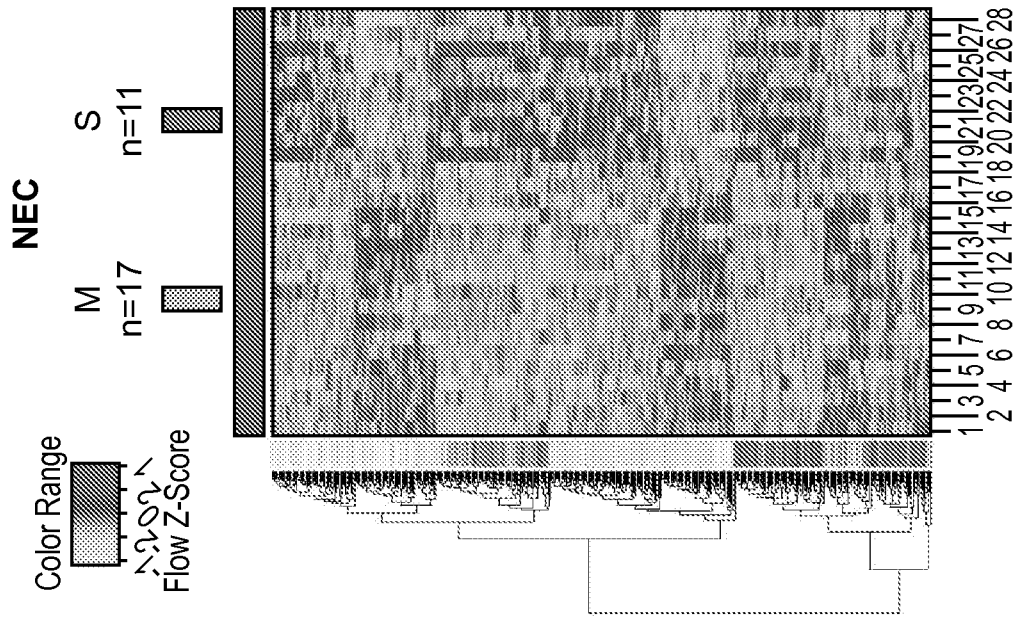

NEC
M
n=17
S
n=11
Color Range
Flow Z-Score
1 3 5 7 9 11 13 15 17 19 21 23 25 27
2 4 6 8 10 12 14 16 18 20 22 24 26 28

FIG. 3A

Protein	Relative abundance M	S	M/Z
COL1A1	-0.2004	0.5635	1425.62
	-0.0481	0.1953	1494.69
	-0.0063	0.0988	1832.86
COL1A2	-0.2323	0.3357	1752.83
	-0.0233	0	1695.79
GLDN	-0.019	0.1002	1669.72
	0	0.0707	1596.81
EMID1	0	0.0275	1886.88
	0	0.0162	1603.76
Teneurin-3	-0.0267	0.0501	1448.41
	-0.0134	0.0426	1336.35
HBA1	-0.1838	0.4524	1874.96
	-0.0933	0.2622	1360.67
SF1	-0.0019	0.2104	2924.4
	0	-0.0257	2918.34
EHBP1L1	0.0161	-0.1421	3013.34
	-0.0518	0.1667	1632.8
UMOD	0.1743	-0.1455	1767.98
	0.1325	-0.0745	2040.15
	0.2168	-0.2198	1680.96
	0.2863	-0.2471	1912.06
COL2A1	0	0.0366	1903.94
	0.0835	-0.098	2058.88
	0	0.1065	2713.18
	0.0686	-0.0076	2324.10
PRAGMIN	0	0.0329	3577.64
	0	-0.1226	3252.49
FGA	-0.1753	0	2659.26
	-0.1735	0.3245	2560.19

FIG. 3C

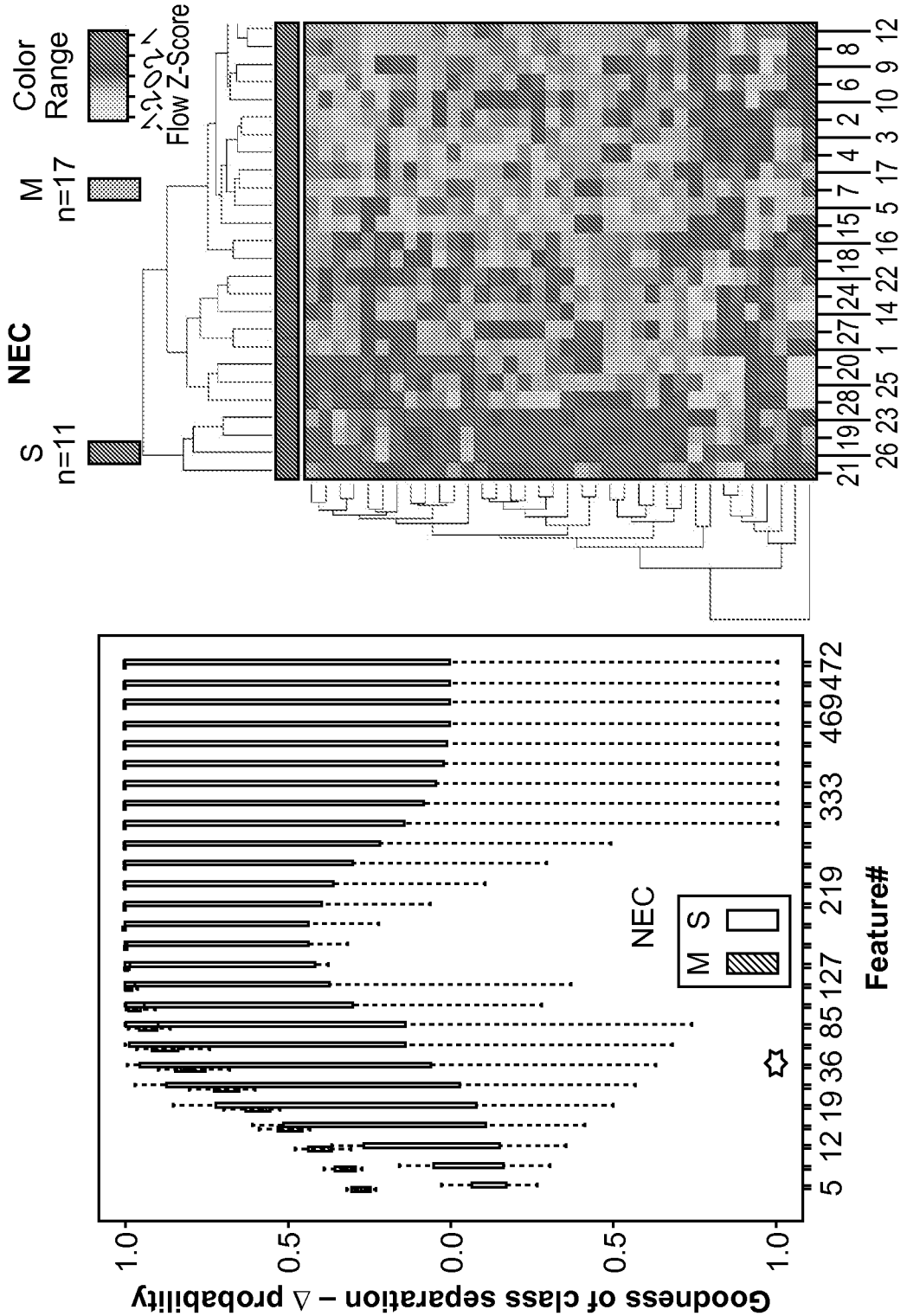

Color Range
Flow Z-Score
M n=17
S n=11
NEC
21 19 28 20 27 24 18 15 7 4 2 6 8
26 23 25 1 14 22 16 5 17 3 10 9 12
Goodness of class separation – Δ probability
1.0 0.5 0.0 0.5 1.0
5 12 19 36 85 127 219 333 469 472
Feature#
NEC
M S

FIG. 4A

FIG. 4B

Heatmap ID	M/Z	Protein	Relative abundance	
			M	S
1	1682.73	SNX15	-0.1619	0.3679
2	1568.74	FGA	-0.115	0.2464
3	1853.82	COL1A2	-0.0639	0
4	1752.83	COL1A2	-0.2323	0.3357
5	2725.28	RASF4	-0.0591	0.0925
6	2088.85	HOXD3	-0.0614	0.1592
7	2578.25	COL18A1	0.1506	-0.1334
8	1296.61	EDA	-0.0919	0
9	2428.09	HUWE1	-0.1949	0.2321
10	1529.74	COL11A2	-0.0084	0.1101
11	1461.72	COL4A2	-0.0278	0.1064
12	2725.4	MUC3A	-0.1125	0.3032
13	1060.51	Q6ZUQ4	-0.0342	0.0041
14	1045.49	MUC15	-0.0424	0.1308
15	1010.47	CRTC1	-0.1197	0.1793
16	1924.93	FETUA	-0.0841	0.1713
17	1077.52	KIAA1012	-0.0365	0.0903
18	1217.58	OBFC2B	0	0.0588
19	1874.96	HBA1	-0.1838	0.4524
20	1143.28	KRTAP5-11	-0.2053	0.2648
21	1305.34	DSG4	-0.1175	0.2859
22	1193.63	COL8A1	-0.0162	0.1193
23	1193.58	FAM83G	-0.0161	0.0136
24	1212.72	GRASP	-0.1286	0.3381
25	1925.99	NBEAL2	-0.1533	0.1225
26	1242.75	YI020	-0.2258	0.3295
27	1268.53	MAML2	-0.1148	0.3077
28	2582.17	HOXD9	0.1591	-0.1422
29	957.028	KV301	0.3129	-0.2552
30	908.971	HV312	0.1254	-0.1414
31	924.929	KV203	0.0732	-0.1711
32	2659.26	FGA	-0.1753	0
33	2560.19	FGA	-0.1735	0.3245
34	1679.8	COL11A2	0	0.0764
35	1680.96	UMOD	0.2168	-0.2198
36	1912.06	UMOD	0.2863	-0.2471

FIG. 4C

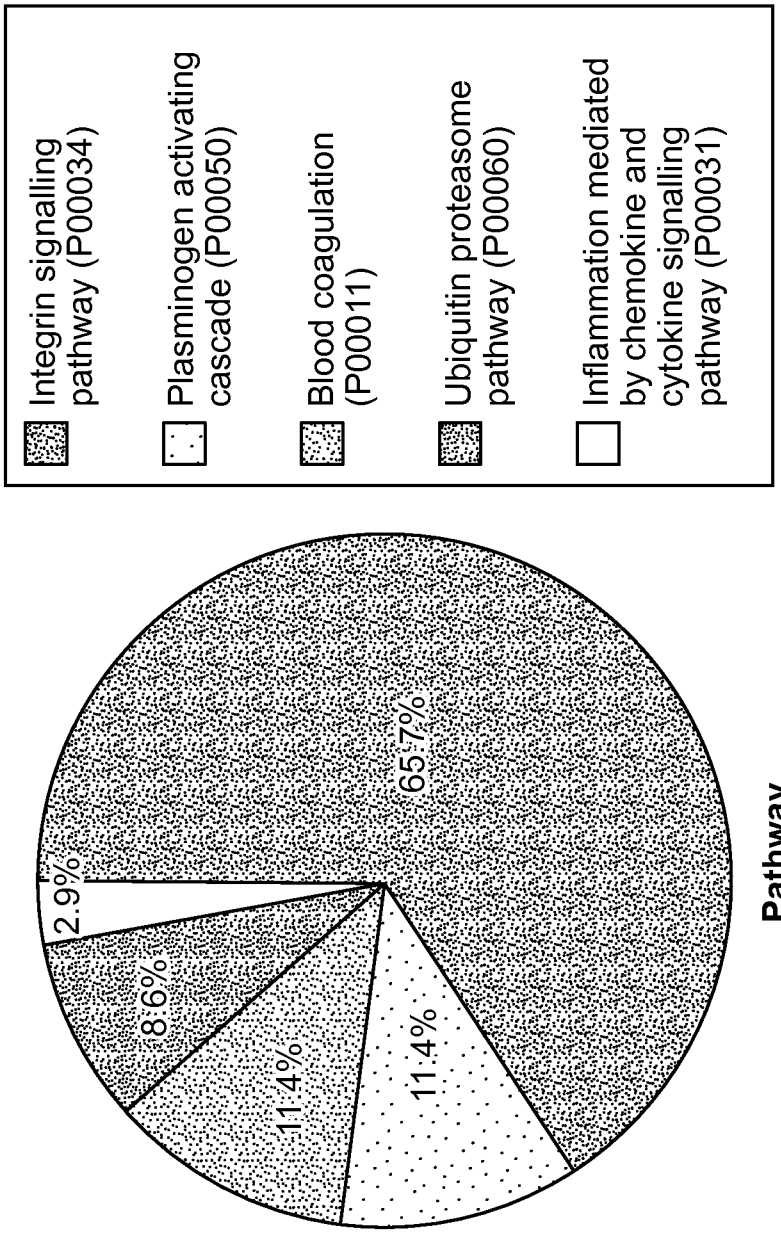

Integrin signalling pathway (P00034)
Plasminogen activating cascade (P00050)
Blood coagulation (P00011)
Ubiquitin proteasome pathway (P00060)
Inflammation mediated by chemokine and cytokine signalling pathway (P00031)
65.7%
11.4%
11.4%
8.6%
2.9%
Pathway

FIG. 4D

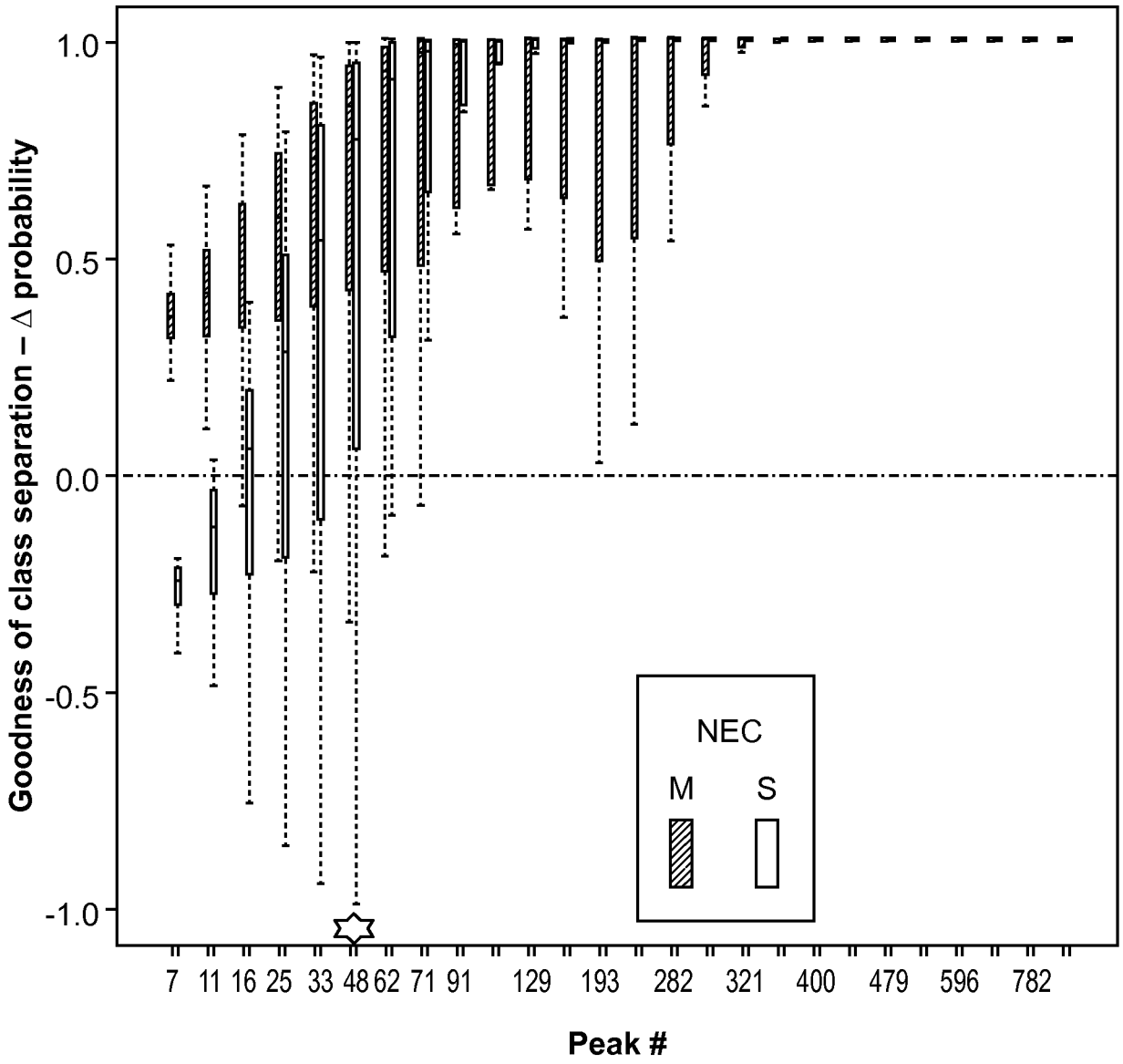

Goodness of class separation – Δ probability
1.0
0.5
0.0
-0.5
-1.0
7 11 16 25 33 48 62 71 91 129 193 282 321 400 479 596 782
Peak #
NEC
M
S

FIG. 5A

Protein ID	MW	Relative abundance		U Test P value	T Test P value
		M	S		
1	2696.00	-0.06	0.13	1.37E-04	1.05E-03
2	3030.00	-0.01	0.02	2.31E-05	7.02E-03
3	3591.00	-0.06	0.12	2.90E-06	1.63E-03
4	4017.00	-0.05	0.10	1.93E-04	2.93E-03
5	5408.00	-0.01	0.01	1.91E-06	1.10E-03
6	6679.00	0.03	-0.06	3.43E-06	2.24E-08
7	7157.50	0.02	-0.04	1.93E-05	4.78E-07
8	7939.00	-0.03	0.06	1.86E-03	3.45E-03
9	7957.00	-0.04	0.07	2.29E-04	1.67E-03
10	8041.50	-0.05	0.09	1.27E-04	8.57E-04
11	8236.33	0.01	-0.02	6.44E-07	1.14E-05
12	8242.00	-0.04	0.07	7.81E-04	1.51E-03
13	8451.00	-0.02	0.05	3.76E-04	2.94E-03
14	8474.50	0.04	-0.08	7.20E-05	1.51E-05
15	8592.75	0.02	-0.04	6.17E-05	4.16E-05
16	8645.50	0.01	-0.03	1.22E-04	7.29E-05
17	8710.00	0.05	-0.09	6.69E-05	4.56E-05
18	8747.00	0.05	-0.11	4.14E-05	2.64E-05
19	8767.00	0.04	-0.08	6.32E-05	4.15E-05
20	8821.00	0.07	-0.14	2.45E-06	5.50E-07
21	8940.00	0.00	0.00	4.70E-05	4.31E-05
22	8955.00	-0.01	0.01	5.58E-02	4.25E-03
23	8978.00	-0.06	0.12	1.71E-02	1.65E-03
24	10296.00	-0.05	0.09	4.04E-06	3.26E-04
25	10858.50	-0.02	0.04	4.95E-06	9.55E-04
26	11006.00	0.00	0.00	1.83E-06	9.63E-04
27	11170.00	-0.03	0.05	6.74E-07	3.06E-03
28	11548.00	-0.02	0.05	4.77E-06	1.98E-03
29	11572.00	-0.02	0.04	4.96E-06	4.35E-04
30	17057.33	0.03	-0.06	5.04E-05	1.34E-05

FIG. 5B

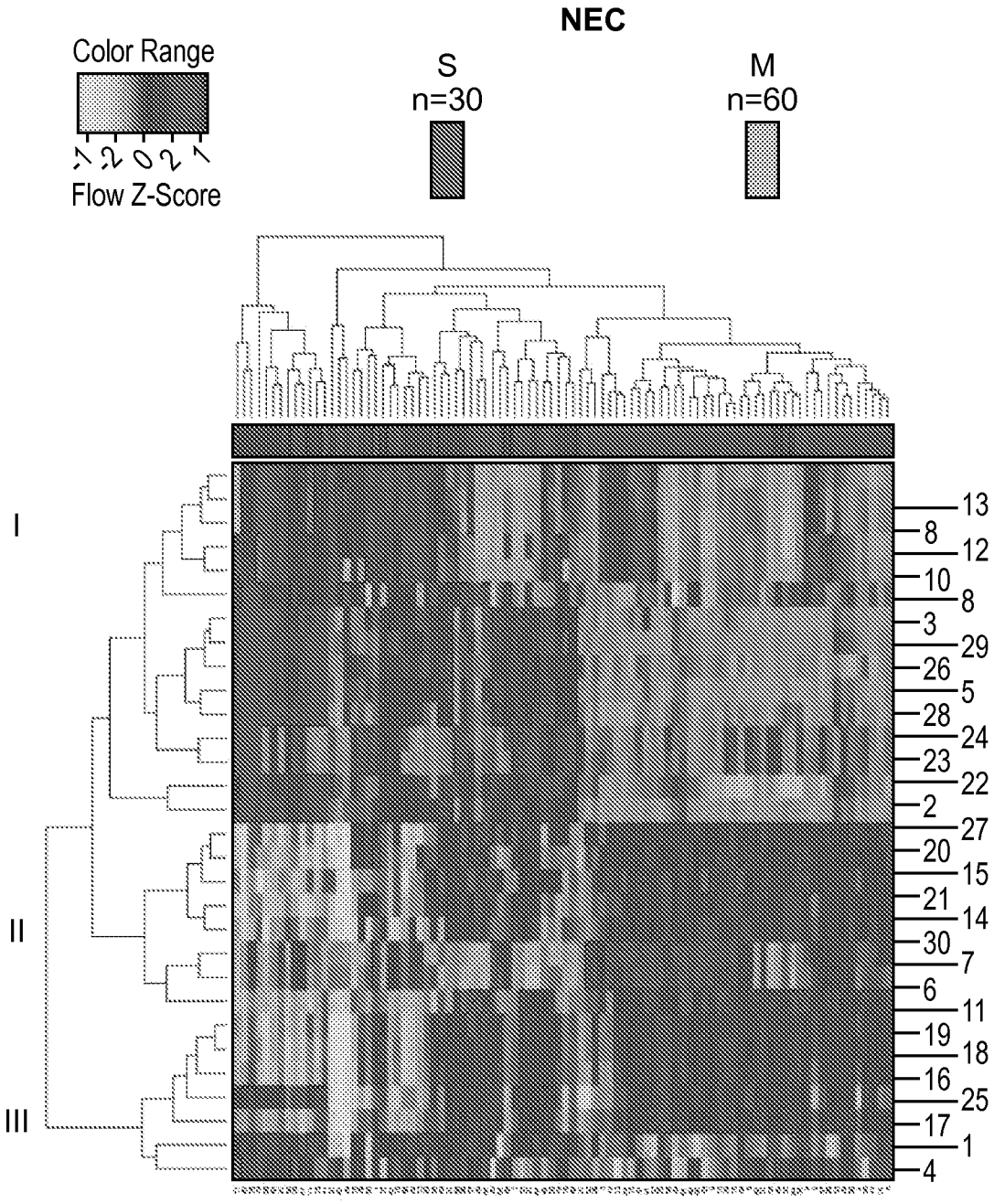

NEC
Color Range
-1 -2 0 2 1
Flow Z-Score
S
n=30
M
n=60
I
II
III
13
8
12
10
8
3
29
26
5
28
24
23
22
2
27
20
15
21
14
30
7
6
11
19
18
16
25
17
1
4

FIG. 5C

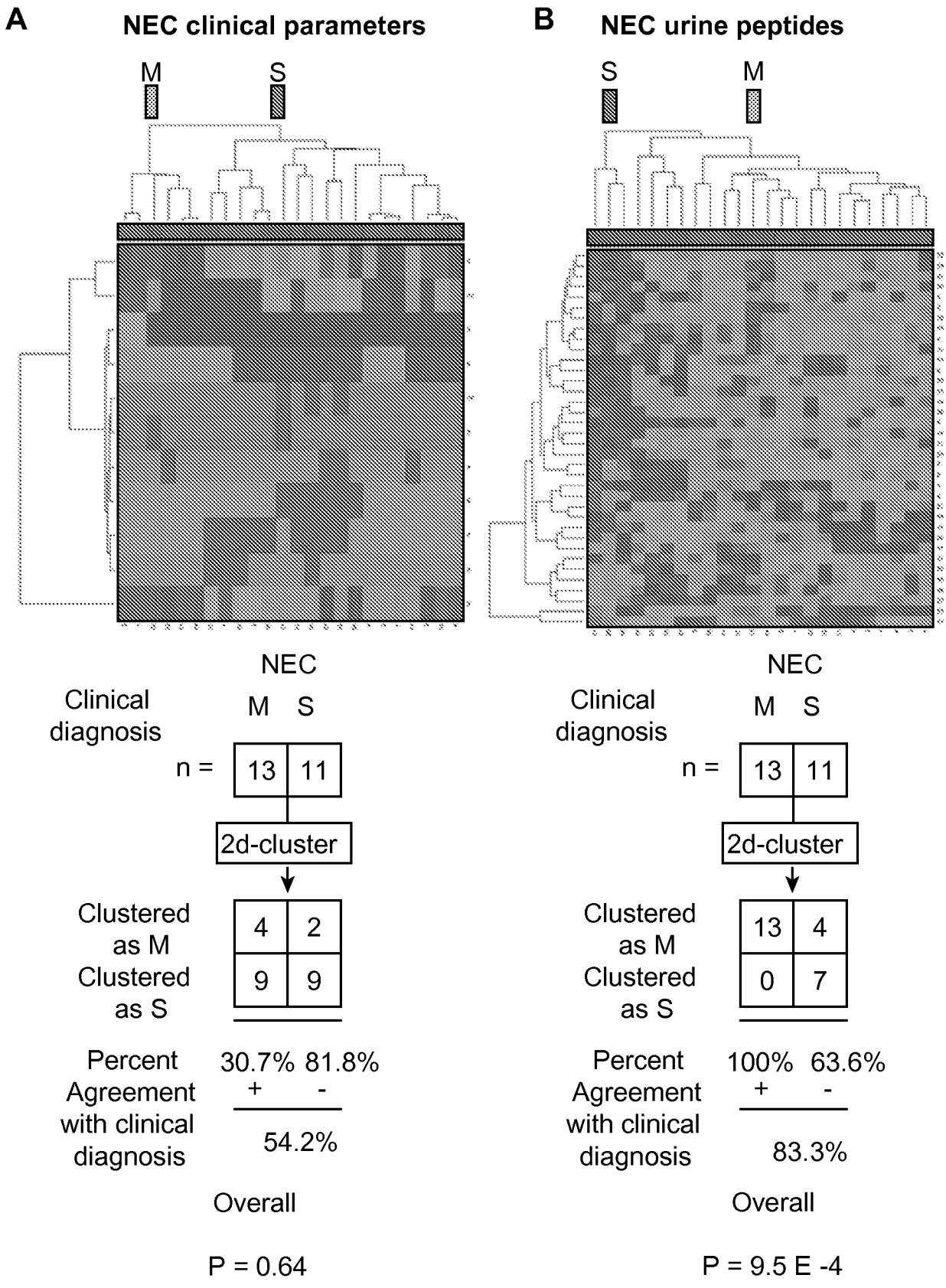

A
NEC clinical parameters
M
S
B
NEC urine peptides
S
M
NEC
Clinical diagnosis
M S
n = 13 11
2d-cluster
Clustered as M
Clustered as S
4 2
9 9
Percent Agreement with clinical diagnosis
30.7% 81.8%
+ -
54.2%
Overall
P = 0.64
NEC
Clinical diagnosis
M S
n = 13 11
2d-cluster
Clustered as M
Clustered as S
13 4
0 7
Percent Agreement with clinical diagnosis
100% 63.6%
+ -
83.3%
Overall
P = 9.5 E -4

FIG. 6

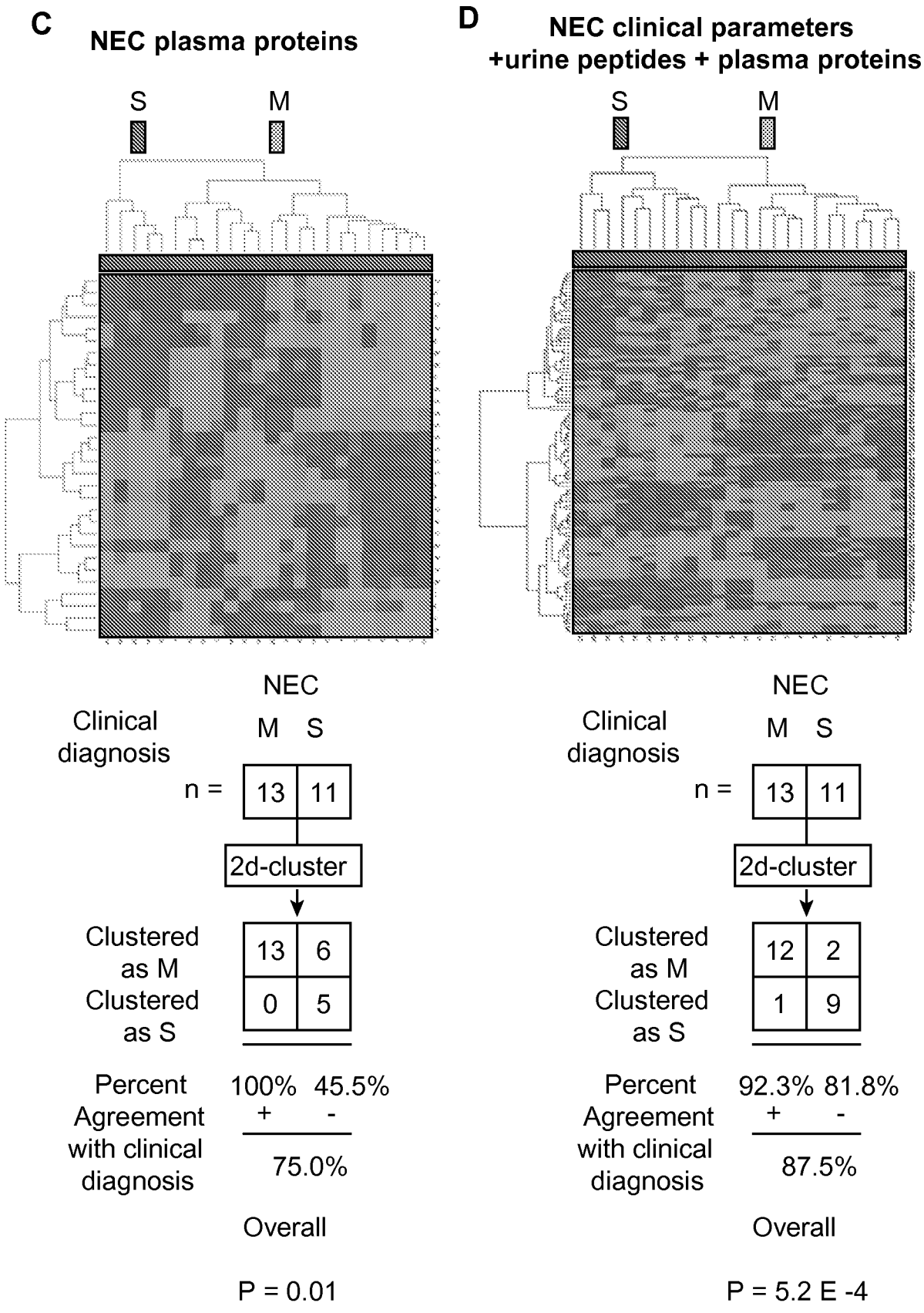

C
NEC plasma proteins
S
M
NEC
Clinical diagnosis
M S
n = 13 11
2d-cluster
Clustered as M
Clustered as S
13 6
0 5
Percent Agreement with clinical diagnosis
100% 45.5%
+ -
75.0%
Overall
P = 0.01
D
NEC clinical parameters
+urine peptides + plasma proteins
S
M
NEC
Clinical diagnosis
M S
n = 13 11
2d-cluster
Clustered as M
Clustered as S
12 2
1 9
Percent Agreement with clinical diagnosis
92.3% 81.8%
+ -
87.5%
Overall
P = 5.2 E -4

FIG. 6 (Cont.)

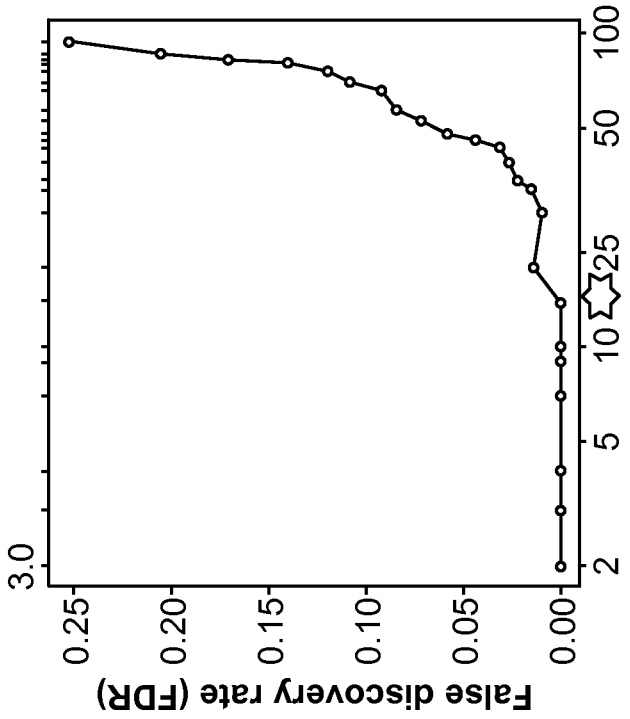

False discovery rate (FDR)
0.25
0.20
0.15
0.10
0.05
0.00
3.0
2
5
10
25
50
100

FIG. 7B

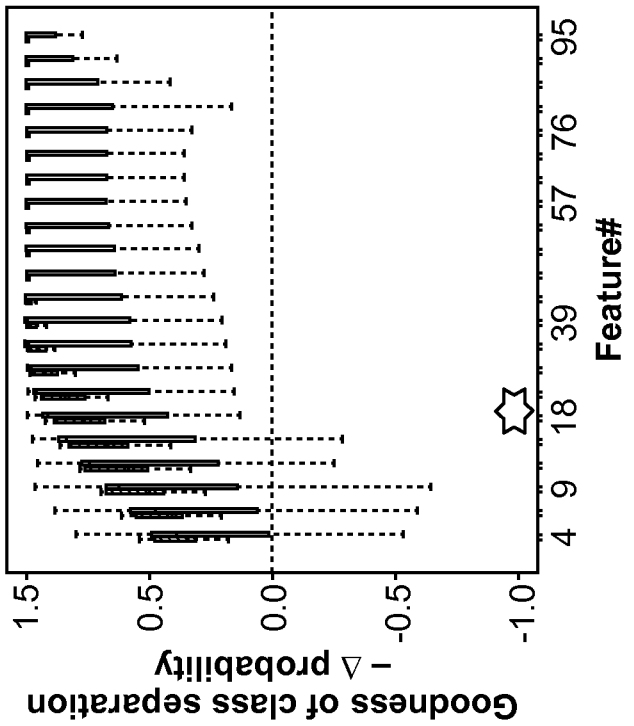
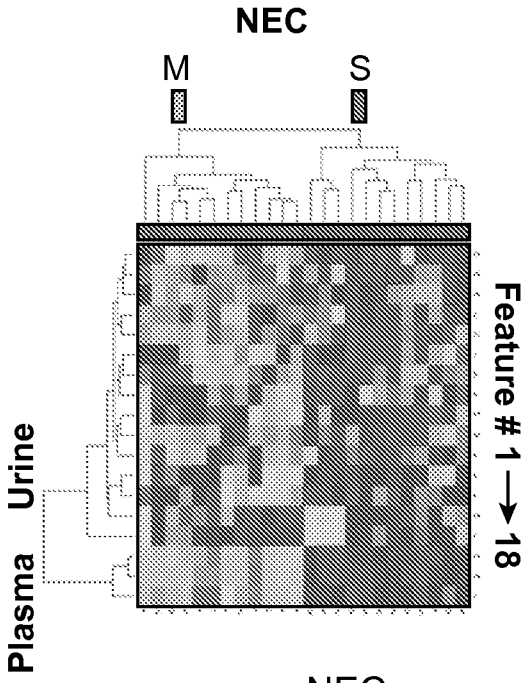
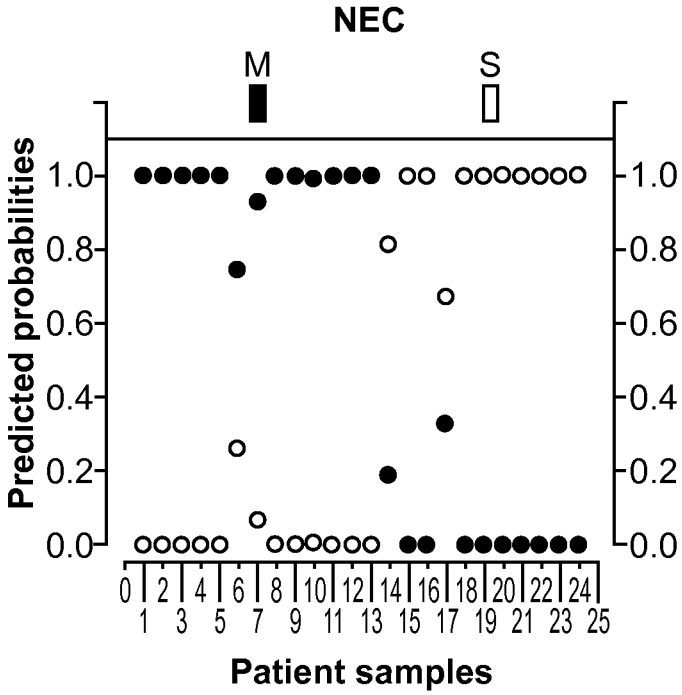

Goodness of class separation
– Δ probability
1.5
0.5
0.0
-0.5
-1.0
4
9
18
39
57
76
95
Feature#

FIG. 7A

	Heatmap index	Marker	NEC		MW	Protein
			M	S		
Urine	1	Urine	-0.04	0.046	1059.51	Q6ZUQ4
	2	Urine	-0.04	0.048	1216.58	OBFC2B
	3	Urine	-0.02	0.024	1528.74	COL11A2
	4	Urine	-0.1	0.119	1924.99	NBEAL2
	5	Urine	-0.02	0.02	1211.72	GRASP
	6	Urine	-0.31	0.363	2427.09	HUWE1
	7	Urine	-0.17	0.201	1751.83	COL1A2
	8	Urine	-0.1	0.116	2087.85	HOXD3
	9	Urine	-0.02	0.026	1304.34	DSG4
	10	Urine	-0.07	0.081	1142.28	KRTAP5-11
	11	Urine	-0.09	0.104	1241.75	YI020
	12	Urine	-0.3	0.356	2658.26	FGA
	13	Urine	-0.08	0.091	2559.18	FGA
	14	Urine	0.033	-0.04	1679.96	UMOD
	15	Urine	0.098	-0.12	1911.06	UMOD
Plasma	16	Plasma	-0.04	0.052	8596.00	
	17	Plasma	-0.02	0.018	8451.00	
	18	Plasma	-0.12	0.147	5408.00	

FIG. 7C


NEC
M
S
Plasma
Urine
Feature # 1 → 18


NEC
M
S
Predicted probabilities
1.0
0.8
0.6
0.4
0.2
0.0
0 2 4 6 8 10 12 14 16 18 20 22 24
1 3 5 7 9 11 13 15 17 19 21 23 25
Patient samples

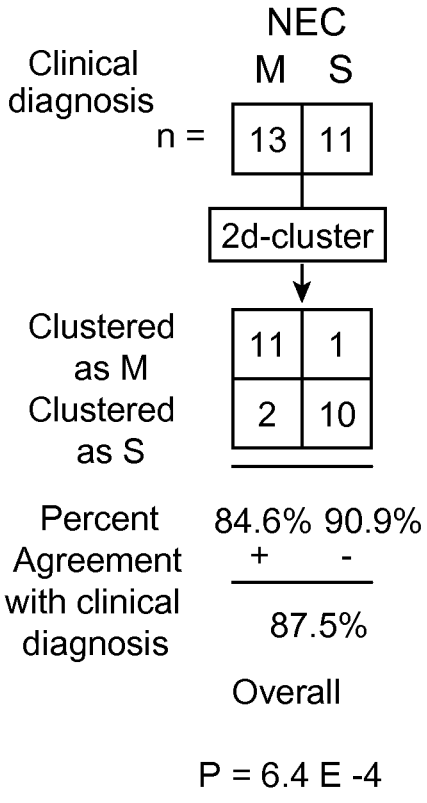

NEC
M S
Clinical diagnosis
n = 13 11
2d-cluster
Clustered as M
Clustered as S
11 1
2 10
Percent Agreement with clinical diagnosis
84.6% 90.9%
+ -
87.5%
Overall
P = 6.4 E -4

FIG. 7D

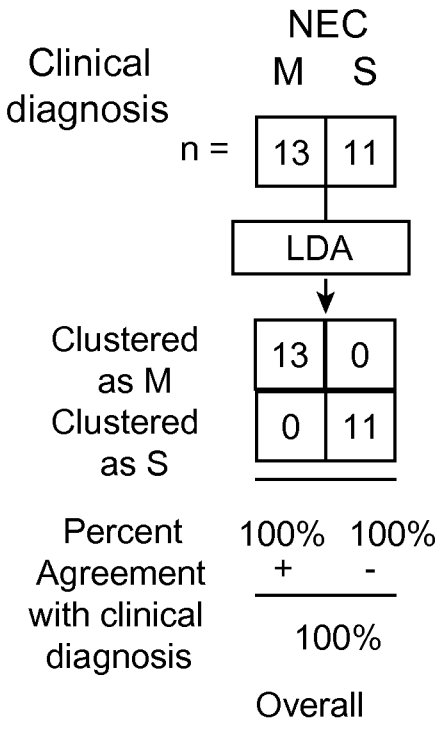

NEC
M S
Clinical diagnosis
n = 13 11
LDA
Clustered as M
Clustered as S
13 0
0 11
Percent Agreement with clinical diagnosis
100% 100%
+ -
100%
Overall
P = 4.0 E -7

FIG. 7E

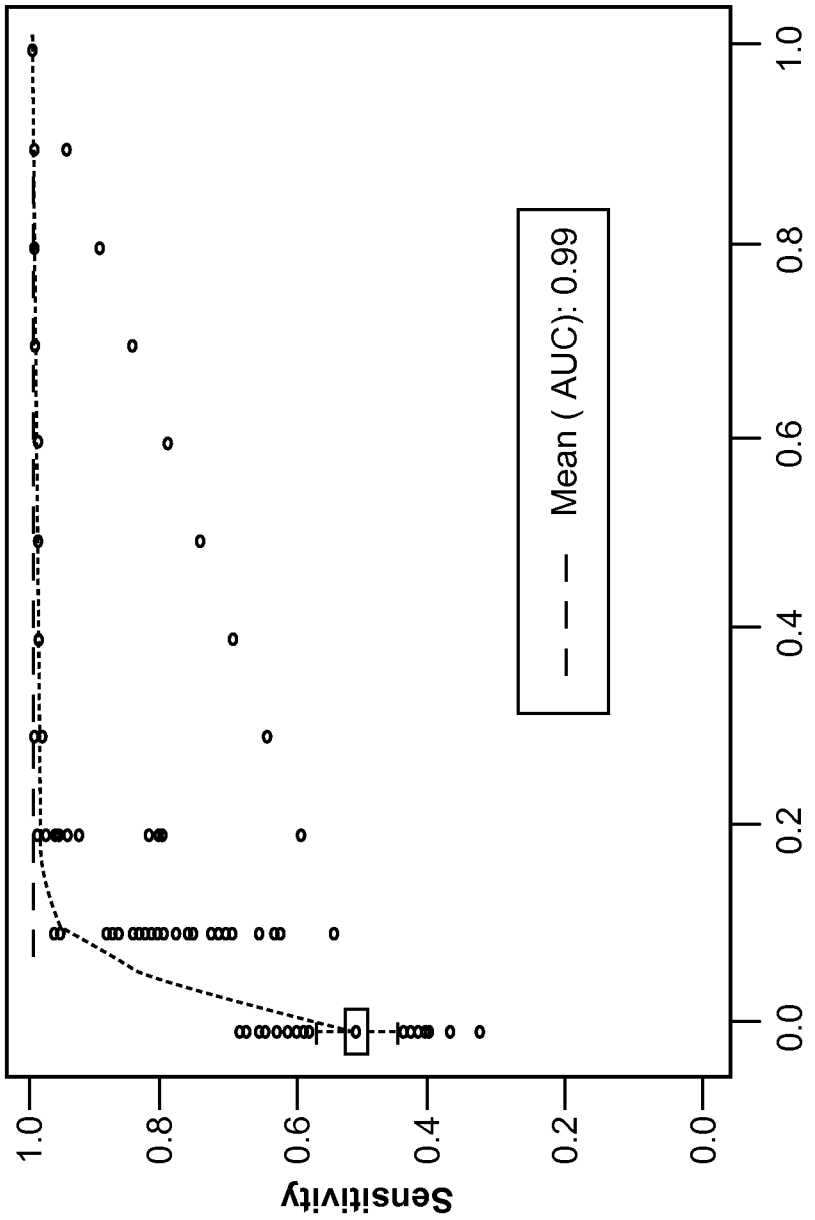
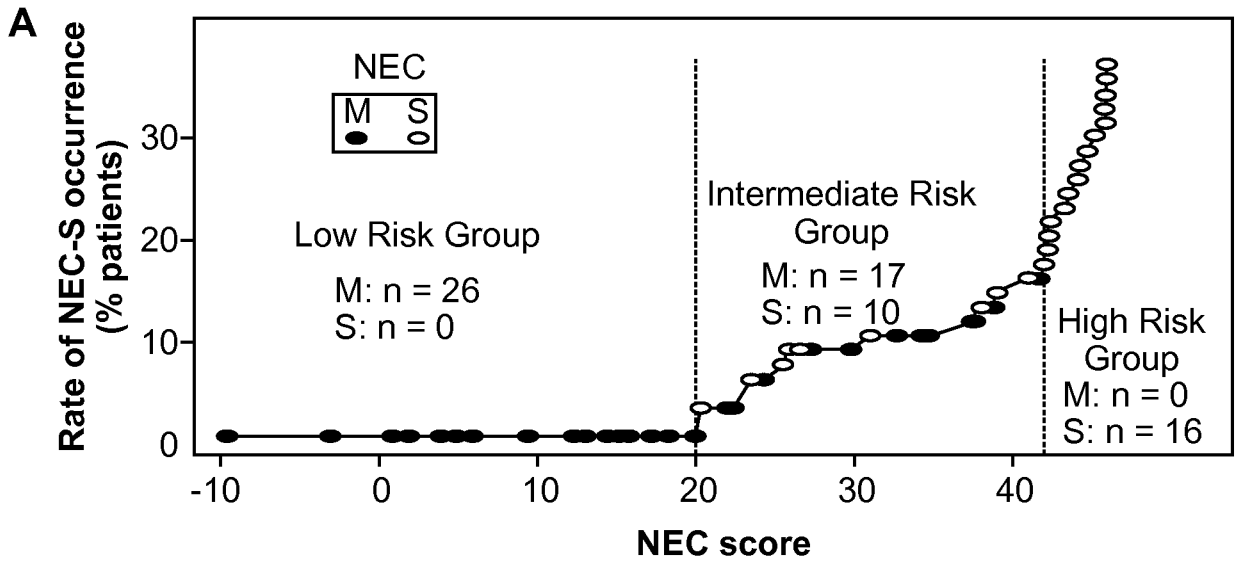

Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
1- Specificity
0.0
0.2
0.4
0.6
0.8
1.0
Mean (AUC): 0.99

FIG. 7F


A
NEC
M S
Rate of NEC-S occurrence (% patients)
30
20
10
0
Low Risk Group
M: n = 26
S: n = 0
Intermediate Risk Group
M: n = 17
S: n = 10
High Risk Group
M: n = 0
S: n = 16
-10
0
10
20
30
40
NEC score

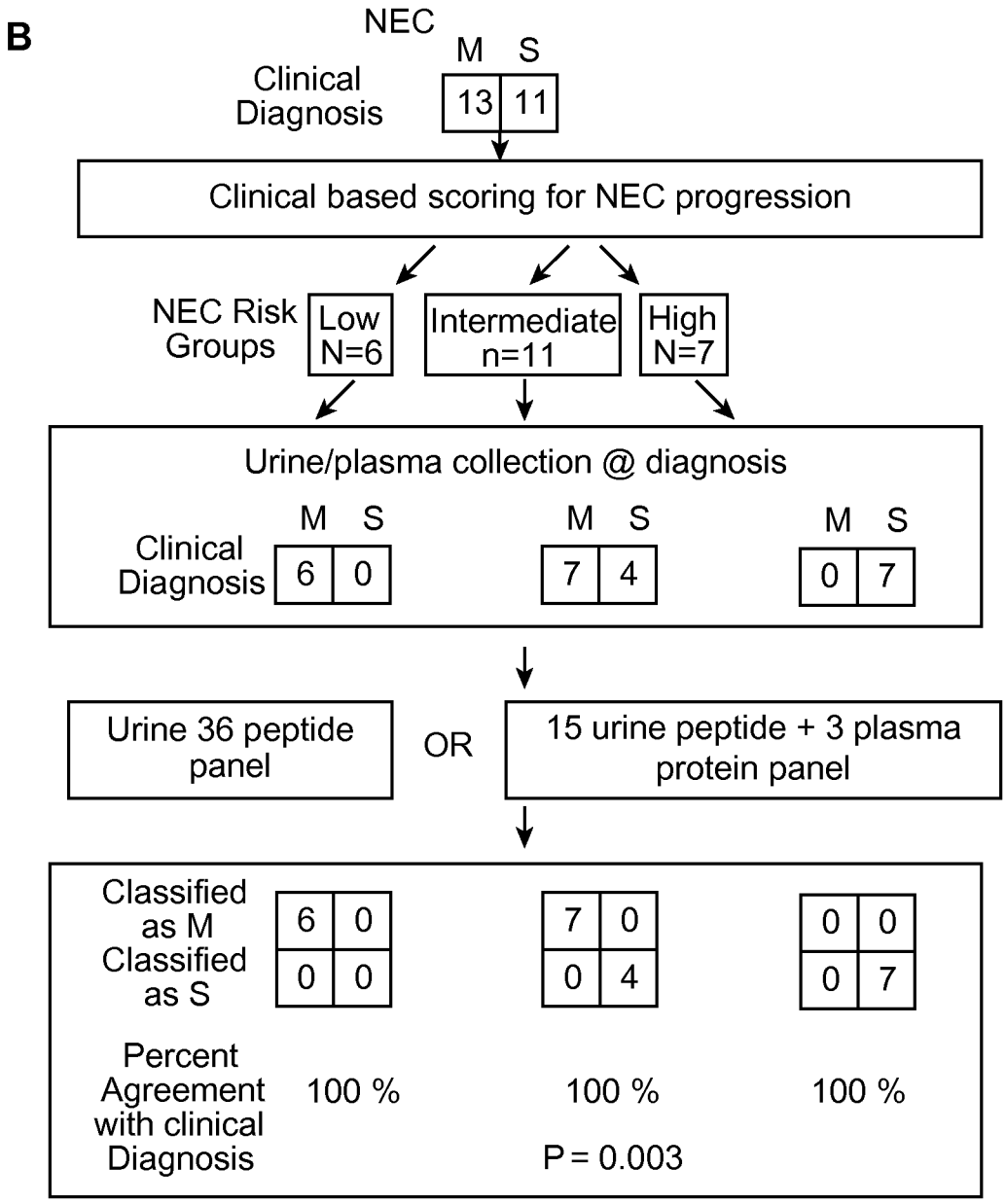

B
NEC
M S
Clinical Diagnosis
13 11
Clinical based scoring for NEC progression
NEC Risk Groups
Low N=6
Intermediate n=11
High N=7
Urine/plasma collection @ diagnosis
M S
Clinical Diagnosis
6 0
M S
7 4
M S
0 7
Urine 36 peptide panel
OR
15 urine peptide + 3 plasma protein panel
Classified as M
Classified as S
6 0
0 0
7 0
0 4
0 0
0 7
Percent Agreement with clinical Diagnosis
100 %
100 %
100 %
P = 0.003

FIG. 8

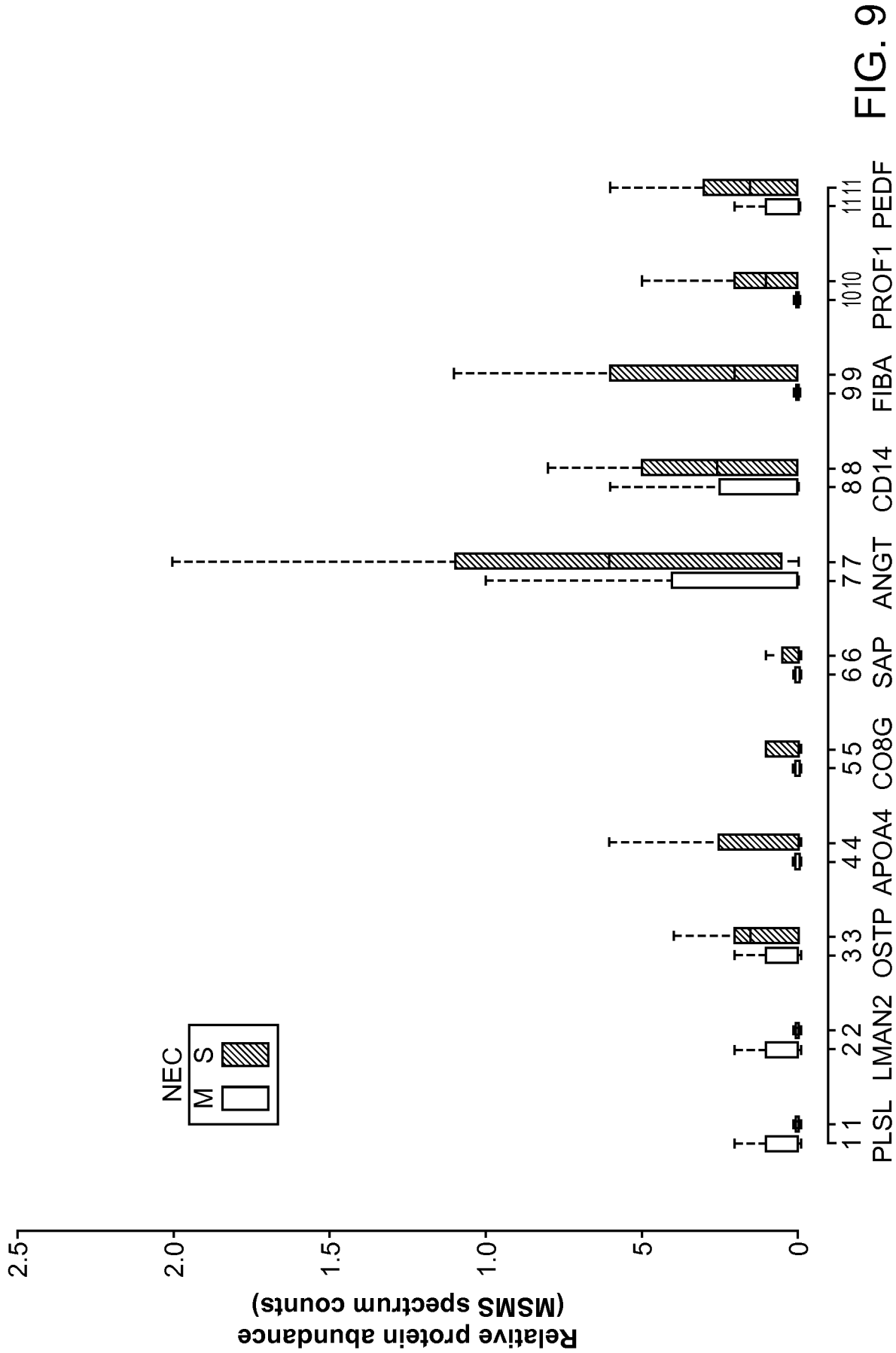

Relative protein abundance
(MSMS spectrum counts)
2.5
2.0
1.5
1.0
5
0
NEC
M S
11 PLSL
22 LMAN2
33 OSTP
44 APOA4
55 CO8G
66 SAP
77 ANGT
88 CD14
99 FIBA
1010 PROF1
1111 PEDF

FIG. 9

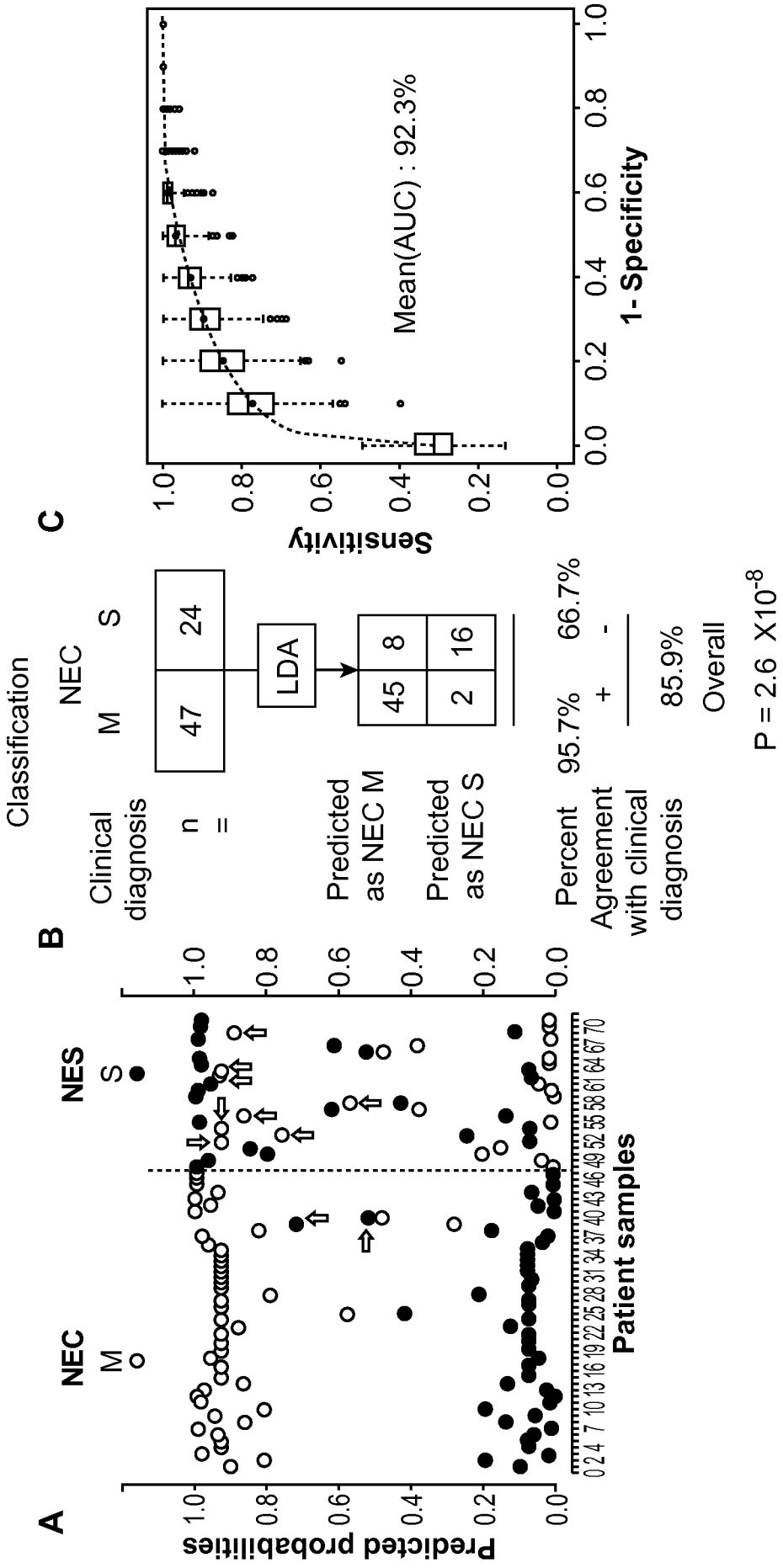

A
NEC
M
NES
S
Predicted probabilities
Patient samples
B
Classification
NEC
M
S
Clinical diagnosis
n =
47
24
LDA
Predicted as NEC M
45
8
Predicted as NEC S
2
16
Percent Agreement with clinical diagnosis
95.7%
66.7%
+
-
85.9%
Overall
P = 2.6 X10-8
C
Sensitivity
1- Specificity
Mean(AUC) : 92.3%

FIG. 10

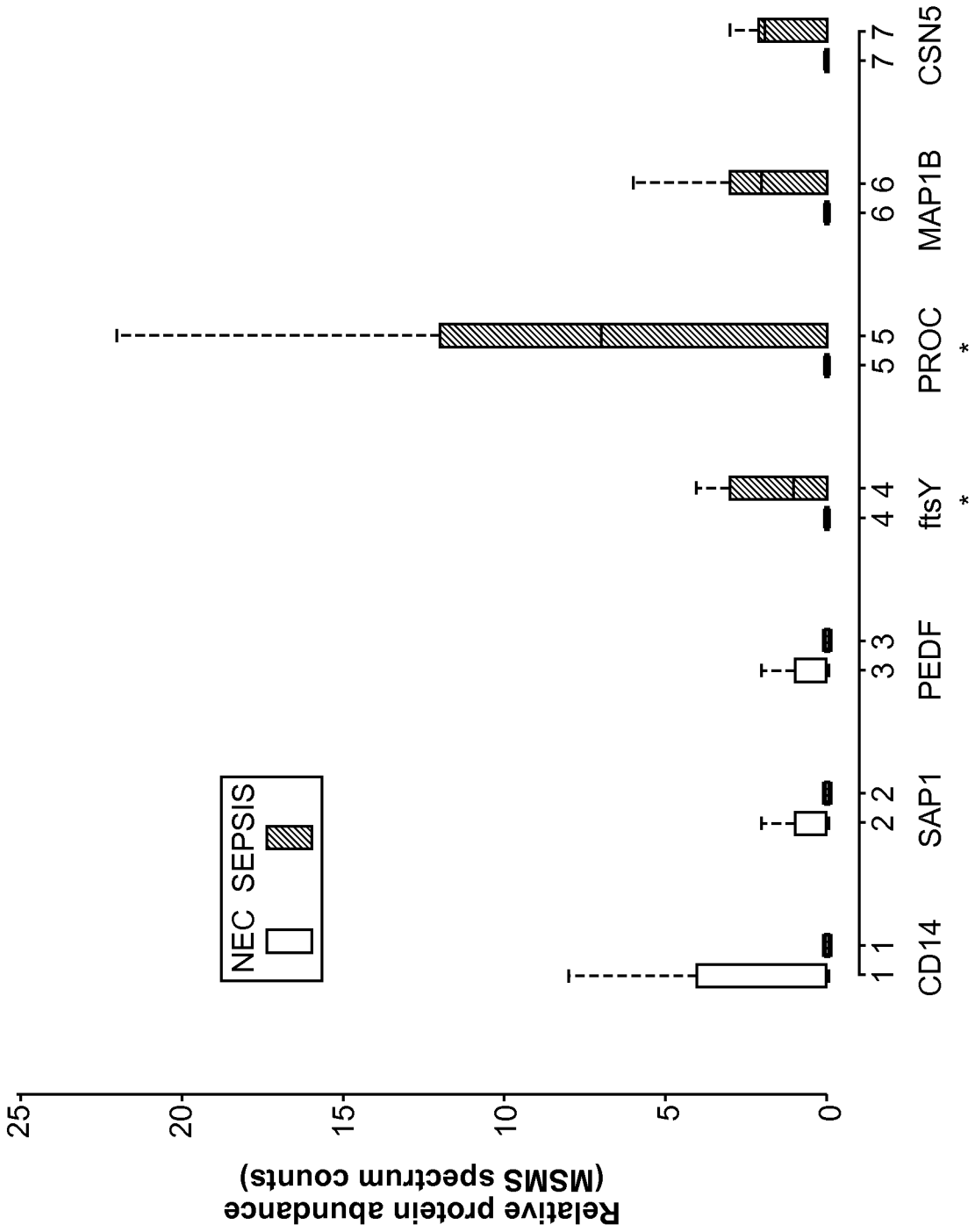

Relative protein abundance
(MSMS spectrum counts)
0
5
10
15
20
25
NEC
SEPSIS
1 1 CD14
2 2 SAP1
3 3 PEDF
4 4 ftsY
5 5 PROC
6 6 MAP1B
7 7 CSN5
*
*

FIG. 11

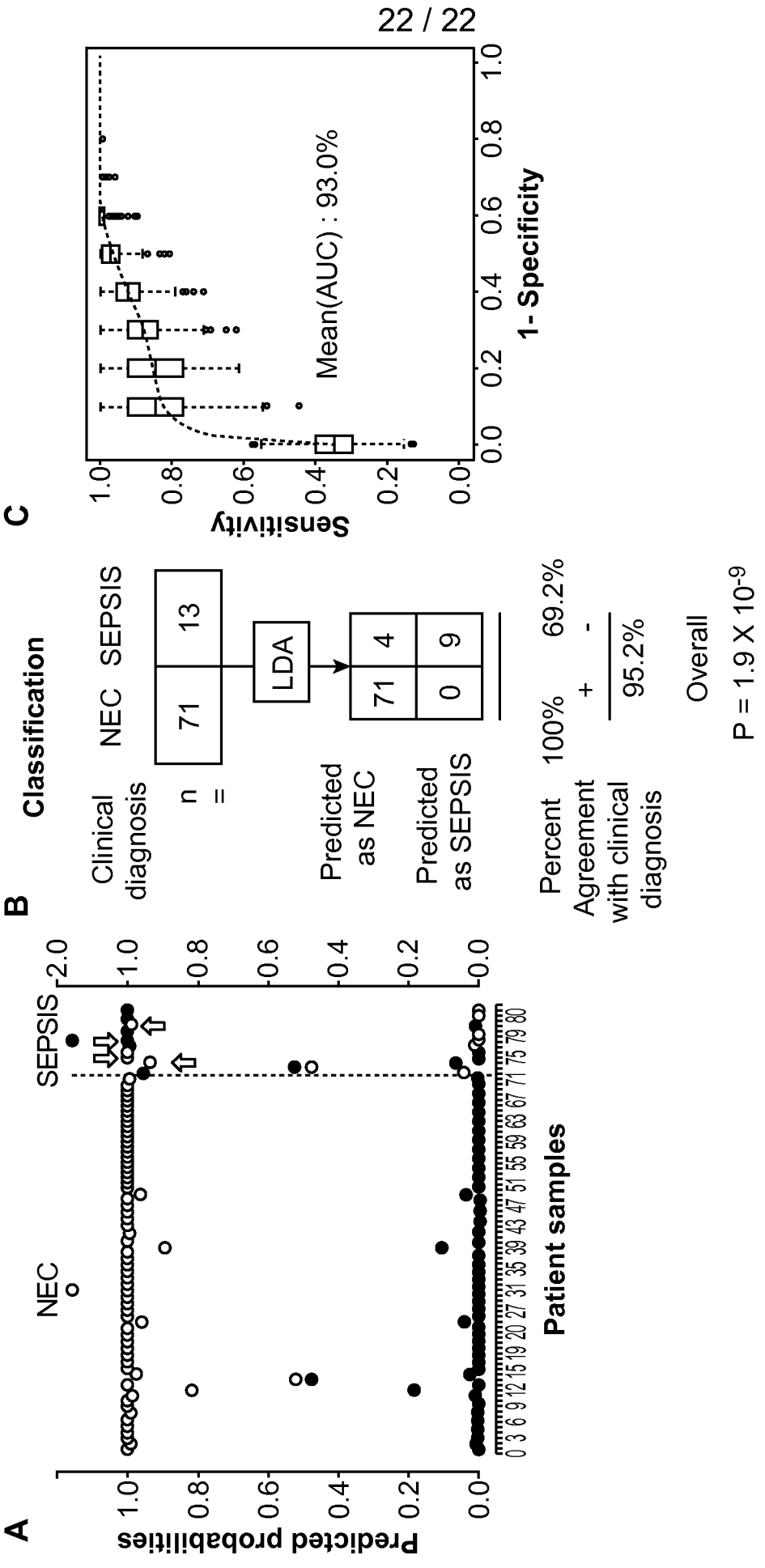

A
Predicted probabilities
1.0
0.8
0.6
0.4
0.2
0.0
NEC
SEPSIS
Patient samples
0 3 6 9 12 15 19 20 27 31 35 39 43 47 51 55 59 63 67 71 75 79 80
B
Classification
Clinical diagnosis
n =
NEC SEPSIS
71 13
LDA
Predicted as NEC
Predicted as SEPSIS
71 4
0 9
Percent Agreement with clinical diagnosis
100% 69.2%
+ -
95.2%
Overall
P = 1.9 X 10-9
C
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
Mean(AUC) : 93.0%
0.0 0.2 0.4 0.6 0.8 1.0
1- Specificity

FIG. 12

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US20 12/042275

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12Q 1/68 (2012.01)
USPC - 435/6.1
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8) C12Q 1/68; G01N 33/50, 33/53 (2012.01)
USPC - 42$_4$/93.4; 435/4, 6.1, 6.18, 7.1; 436/518

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PatBase, Google Patent, Google, Google Scholar, Pubmed

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 7,871,785 B2 (MORROW et al) 18 January 2011 (18.01.2011) entire document	1-13
Y	US 2010/0008925 A1 (LUK et al) 14 January 2010 (14.01.2010) entire document	1-4, 8-13
Y	US 2009/0163599 A1 (STORR et al) 25 June 2009 (25.06.2009) entire document	5-7
Y	US 6,804,551 B2 (GRIFFIN et al) 12 October 2004 (12.10.2004) entire document	3, 13
Y	LING et al. Urine Peptidomics for Clinical Biomarker Discovery. Advances in Clinical Chemistry. Volume 51. 2010. Pages 181-210.	10, 11

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search
20 August 2012

Date of mailing of the international search report
27 SEP 2012

Name and mailing address of the ISA/US
Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450
Facsimile No. 571-273-3201

Authorized officer:
Blaine R. Copenheaver
PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)