



WAGENINGEN  
UNIVERSITY & RESEARCH

## Systems Biology and Biomarkers in Necrotizing Soft Tissue Infections

Necrotizing Soft Tissue Infections

Saccenti, Edoardo; Svensson, Mattias

[https://doi.org/10.1007/978-3-030-57616-5\\_11](https://doi.org/10.1007/978-3-030-57616-5_11)

This article is made publicly available in the institutional repository of Wageningen University and Research, under the terms of article 25fa of the Dutch Copyright Act, also known as the Amendment Taverne. This has been done with explicit consent by the author.

Article 25fa states that the author of a short scientific work funded either wholly or partially by Dutch public funds is entitled to make that work publicly available for no consideration following a reasonable period of time after the work was first published, provided that clear reference is made to the source of the first publication of the work.

This publication is distributed under The Association of Universities in the Netherlands (VSNU) 'Article 25fa implementation' project. In this project research outputs of researchers employed by Dutch Universities that comply with the legal requirements of Article 25fa of the Dutch Copyright Act are distributed online and free of cost or other barriers in institutional repositories. Research outputs are distributed six months after their first online publication in the original published version and with proper attribution to the source of the original publication.

You are permitted to download and use the publication for personal purposes. All rights remain with the author(s) and / or copyright owner(s) of this work. Any use of the publication or parts of it other than authorised under article 25fa of the Dutch Copyright act is prohibited. Wageningen University & Research and the author(s) of this publication shall not be held responsible or liable for any damages resulting from your (re)use of this publication.

For questions regarding the public availability of this article please contact [openscience.library@wur.nl](mailto:openscience.library@wur.nl)



# Systems Biology and Biomarkers in Necrotizing Soft Tissue Infections

# 11

Edoardo Saccenti and Mattias Svensson

## Contents

11.1	<b>Introduction</b> .....	168
11.2	<b>State of the Art: Existing Biomarkers for NSTI</b> .....	169
11.3	<b>Biomarker Type and Detection</b> .....	169
11.3.1	Diagnostic Biomarkers .....	170
11.3.2	Prognostic Biomarkers .....	170
11.3.3	Predictive Biomarkers .....	171
11.3.4	Predisposing Biomarkers .....	171
11.4	<b>Biomarker Use</b> .....	172
11.4.1	Genomic Biomarkers in NSTI .....	172
11.4.2	Transcriptomic Biomarkers in NSTI .....	172
11.4.3	Epigenomic Biomarkers in NSTI .....	173
11.4.4	Proteomic Biomarkers in NSTI .....	173
11.4.5	Metabolomic Biomarkers in NSTI .....	174
11.4.6	Microbiomic Biomarkers in NSTI .....	174
11.4.7	From One-Category-Analyses Towards Systems Biological Approaches ...	174
11.5	<b>Systems Biology</b> .....	175
11.6	<b>Systems Biology Approaches to Biomarker Discovery</b> .....	176
11.7	<b>Networks and Network Analysis</b> .....	177
11.8	<b>Multivariate Biomarker and Omics Signatures</b> .....	180
11.9	<b>Future Perspective</b> .....	182
	<b>References</b> .....	184

E. Saccenti (✉)  
Laboratory of Systems and Synthetic Biology,  
Wageningen University and Research, Wageningen, The  
Netherlands  
e-mail: [edoardo.saccenti@wur.nl](mailto:edoardo.saccenti@wur.nl)

M. Svensson  
Center for Infectious Medicine, Department of Medicine,  
ANA Futura, Karolinska Institutet, Karolinska University  
Hospital, Huddinge, Sweden

## Abstract

In necrotizing soft tissue infection (NSTI) there is a need to identify biomarker sets that can be used for diagnosis and disease management. The INFECT study was designed to obtain such insights through the integration of patient data and results from different

clinically relevant experimental models by use of systems biology approaches. This chapter describes the current state of biomarkers in NSTI and how biomarkers are categorized. We introduce the fundamentals of top-down systems biology approaches including analysis tools and we review the use of current methods and systems biology approaches to biomarker discover. Further, we discuss how different “omics” signatures (gene expression, protein, and metabolites) from NSTI patient samples can be used to identify key host and pathogen factors involved in the onset and development of infection, as well as exploring associations to disease outcomes.

### Keywords

Epigenomics · Genomics · Network inference and analysis · Mathematical modelling · Metabolomics · Microbiome · Multivariate statistics · Proteomics · Top-down approaches

### Highlights

- Comprehensive systems biology analyses and interpretation of omics data are emerging as valuable tools for biomarker discovery in NSTI.
- The inflammatory proteins IL1 $\beta$ , IL-6, CXCL9, CXCL10, and CXCL11, were identified as potential biomarkers to discriminate *S. pyogenes* NSTI from NSTI caused by other pathogens.
- Metabolite–metabolite association networks analysis has proven able to define sets of metabolites associated with biofilm formation in NSTI.
- Metabolites in combination with multivariate and machine learning modelling have the potential to be utilized as predictive markers in NSTI.

## 11.1 Introduction

In search for biomarkers in NSTI, analyses have usually focused specifically on measuring certain selected factors, such as cytokines. This approach is, to some extent, distorted because it is based on already existing knowledge that somehow

indicates that these factors must be associated with the disease. In this context, a systems biological approach, used correctly, has the greater opportunity to identify hitherto unknown factors that may be associated with disease and that may be used as biomarkers.

Systems biology is a holistic approach to deciphering the complexity of biological systems that starts from the understanding that the networks that form the whole of living organisms are more than the sum of their parts. The systems biology approach to understand disease-related biology has the potential to revolutionize our understanding of the cellular pathways and gene networks underlying the onset of disease, and the mechanisms of treatment strategies that ameliorate disease phenotypes. In the post-genomic era, a wealth of novel approaches for generating and analyzing large, high-dimensional genomic, transcriptomic, proteomic, and metabolomic data sets from cohorts of normal and diseased individuals can be used in combination with systems biology approaches (Rosato et al. 2018). In research related to infectious diseases, such as NSTI, the data sets include information from both pathogen- and host-oriented analyses that can be used at a systems biology level. The emerging field of systems biology attempts to harness complex, multi-parameter systems by computationally integrating gene-level data with molecular pathways and networks to extract new biological insight. Systems biology may combine and augment current strategies to biomarker discovery, generating novel, experimentally testable candidates.

A biomarker is a measurable indicator of some biological state or condition, and in recent years systems biology approaches have been applied to screen and identify diagnostic, prognostic, predictive, and predisposing biomarkers as well as targets for prevention and treatment of disease (Feala et al. 2013; Vafaei et al. 2018; Anvar et al. 2018). There are several successful examples of molecular biomarkers that are currently the clinical standard for diagnostic screening in several diseases, for example, in myocardial infarction (Hajar 2016) and certain cancers (Mehta et al. 2010; Goossens et al. 2015). The search for novel molecular biomarkers

continues to be a major research drive in many biomedical fields. A PubMed (<https://pubmed.ncbi.nlm.nih.gov/>) query for NSTI and biomarker (Biomarker AND (NSTI OR “necrotising soft tissue infection” OR “necrotising fasciitis”) in March 2020 returned only 39 hits (44 using the spelling “necrotizing”) demonstrating how little is known and how limited the research activity on this topic is, despite an incidence ranging from 0.3 to 15.5 case per 100,000 population (Stevens and Bryant 2017) and all-cause 90-day mortality of 18% (Madsen et al. 2019). In NSTI biomarkers for early detection and guidance of treatment may improve outcome and help to reduce the length of hospital stay and mortality. There are good opportunities that systems biology approaches in NSTI can identify biomarkers that could be used as novel tools for rapid and early identification and stratification of patients with NSTI upon admission. In this book chapter, we review biomarker types and detection methods, as well as a number of systems biology approaches potentially useful to identify candidate biomarkers in NSTI.

---

## 11.2 State of the Art: Existing Biomarkers for NSTI

There are currently no approved biomarkers for the diagnosis or prognosis in NSTI, and the molecular mechanisms of host responses during NSTI remain poorly understood. This lack of understanding reflects the complex, multifactorial nature of NSTI, which is believed to involve a network of interweaving host and pathogen molecular pathways that dictate disease onset and progression.

Available information on NSTI biomarkers are restricted to a few research studies that focused on a limited number of selected factors or information of factors included as components of relatively inaccurate scoring systems, such as the LRINEC (Laboratory Risk Indicator for Necrotizing Fasciitis). LRINEC is generated from six routinely performed laboratory tests including the analyses of patients' C-reactive protein, white blood cell count ( $\times 10,000/\mu\text{L}$ ), hemoglobin (g/dL), sodium (mEq/L), creatinine, and glucose.

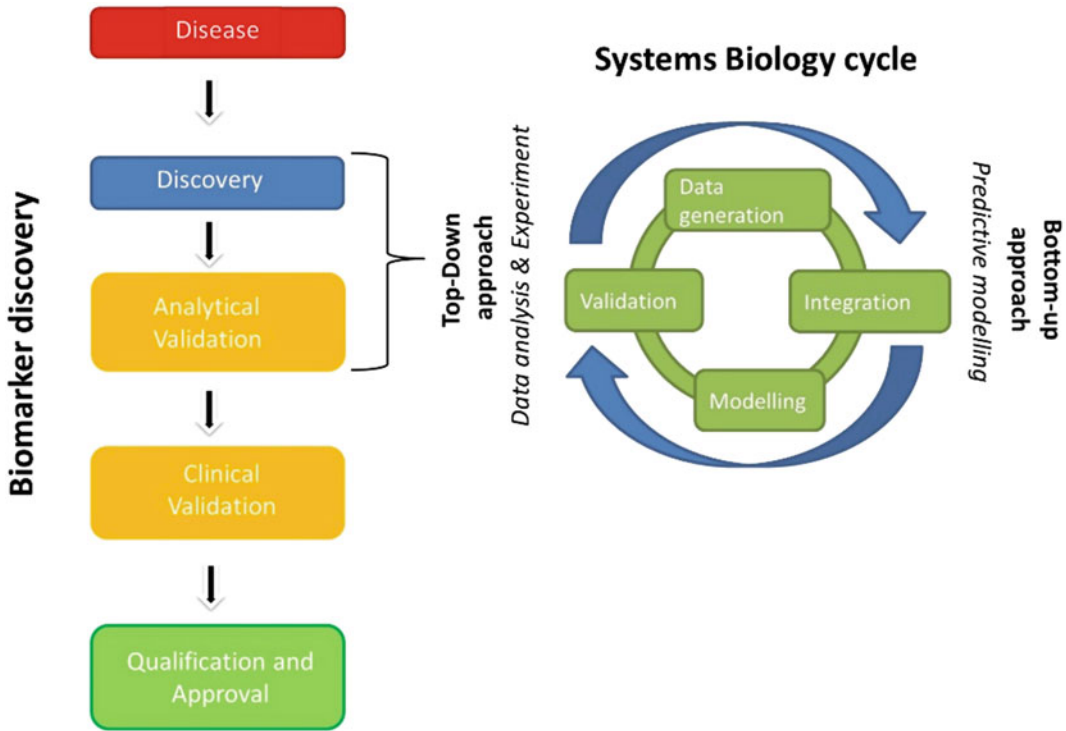
Although LRINEC scoring is not optimal, it has been used to distinguish NSTI from less severe soft tissue infections (Wong et al. 2004).

Recently, LRINEC score has also been analyzed in relation to pro-inflammatory cytokines, and this revealed no significant association between the LRINEC score and analyzed specific cytokine levels on admission (Hansen et al. 2017a). However, the analyses of cytokines revealed IL-6 to be associated with disease severity, and that IL-1 $\beta$  was strongly associated with 30-day mortality (Hansen et al. 2017b). Another factor suggested to have biomarker potential in NSTI is Pentraxin-3 (PTX3) a glycoprotein which belongs to the pentraxin protein superfamily, including acute phase proteins, C-reactive protein (CRP), and serum amyloid protein (SAP). PTX3 is produced by vascular cells or inflammatory cells and is released into the circulation, possibly reflecting systemic inflammation. The PTX3 level in patients with NSTI at time of admission was associated with septic shock, amputation, and risk of death in patients, but was not an independent predictor of 180-day mortality in this patient group (Hansen et al. 2016a). Levels of s-lactate  $>6$  mM and creatinine  $>144$  mM are considered prognostic markers. In addition, microbiological cultures of *Clostridium* or *Vibrio* are negatively associated, while a microbiological culture of *GAS* is positively associated with survival in NSTI. Nevertheless, there is a great need for robust early measurable biomarkers to improve identification and management of NSTI.

---

## 11.3 Biomarker Type and Detection

Molecular biomarkers generally consist of molecules measured in body fluids or in samples from the affected tissue. Challenges facing biomarker research include the lack of standardized methods for fast reliable detection. Biomarkers can be divided into four categories: diagnostic, prognostic, predictive, and predisposing/susceptibility (Quezada et al. 2017; Simon 2011, 2014; Drucker and Krapfenbauer 2013). The typical pattern of biomarker discovery is shown in Fig. 11.1.



**Fig. 11.1** Relationship between the classical biomarker discovery pipeline and the systems biology cycle. Systems biology can play a pivotal role in both discovery and analytical validations steps. Figure adapted from Del Campo et al. (2015) and Rosato et al. (2018)

### 11.3.1 Diagnostic Biomarkers

Diagnostic biomarkers are used, e.g. to determine the health status of a patient to ensure appropriate management. In infectious diseases such as NSTI and sepsis, immediate treatment is required, necessitating quick, early, and accurate diagnosis which can help decision making. A singular ideal biomarker may not be identified, but rather an alternative approach focusing to determine the diagnostic relevancy of multiple biomarkers when used in concert may be useful. The ongoing efforts in the development of a multiplex point-of-care testing kit, enabling quick and reliable detection of serum biomarkers, may have great potential for early diagnosis of acute bacterial infections. Thänert et al. (2019) provided insight into the pathophysiology of mono- and polymicrobial NSTIs and protein multiplex analysis highlighted three potential biomarkers, CXCL9, CXCL10, CXCL11, as these three

chemokines displayed statistically significant concentration differences between *S. pyogenes* and polymicrobial NSTIs. In addition, this study performed RNA-seq analyses on tissue biopsies and this showed that genes with significantly greater expression in polymicrobial compared to monomicrobial NSTIs were those encoding extra cellular matrix (ECM) components like collagen, fibronectin, lumican, and connective tissue growth factor. In addition, gene sequencing data showed, in line with protein data, that genes encoding interferon-inducible mediators such as *CXCL9*, *CXCL10*, *CXCL11* were expressed at higher levels in monomicrobial NSTIs, in particular those caused by *S. pyogenes*.

### 11.3.2 Prognostic Biomarkers

A prognostic marker identifies outcome in patients. Usually, these markers provide insight

into unfavorable outcome, such as mortality. The presence or absence of a prognostic marker can be useful for the selection of patients for a specific treatment but does not directly predict the response to a treatment.

In NSTI plasma biomarkers representing the early inflammatory response are thought to be useful as prognostic markers of disease severity and mortality. To determine the association of admission biochemical markers, such as lactate, to limb loss and mortality both univariate and multivariable analyses have been used. For example, arterial lactate was found to be associated with both mortality (odds ratio [OR], 1.5; 95% confidence interval [CI], 1.1 to 2.0;  $P$ -value = 0.009) and limb loss (OR, 1.3; 95% CI, 1.0 to 1.7;  $P$ -value = 0.02), and lactate was suggested to be used early on to guide aggressive therapeutic interventions (Schwartz et al. 2013). In addition, attempts have been made to associate markers with ICU (Intensive Care Unit) scoring systems including simplified acute physiology score (SAPS) II (Le Gall et al. 1993) and sequential organ failure assessment (SOFA) scores (Vincent et al. 1996), as well as the LRINEC score (Wong et al. 2004), presence of septic shock, microbial etiology, renal replacement therapy, and amputation.

In a study by Hansen et al. it was demonstrated that plasma levels of mannose-binding lectin and ficolins, which can activate the complement pathway, were significantly lower in patients with NSTI than in controls (Hansen et al. 2016b). Furthermore, this study revealed that a high baseline ficolin-2 level indicated a 94% chance of surviving the first 28 days after admission. Another study by Hansen et al. reported on the association between cytokine (IL-1 $\beta$ , IL-6, IL-10, and TNF- $\alpha$ ) levels and the LRINEC score, disease severity and mortality in NSTI patients (Hansen et al. 2017b). Although no significant association between the LRINEC score and cytokine levels on admission was found, the authors found that IL-6 was consistently associated with disease severity and that high IL-1 $\beta$  (OR 3.86 [95% CI, 1.43–10.40],  $P$ -value = 0.008) and IL-10 (4.80 [1.67–13.78],  $P$ -value = 0.004) were associated with 30-day mortality. Also,

PTX3 may be an appropriate marker of severity and prognosis of NSTI (Hansen et al. 2016a). Although PTX3 was reported to perform even better than the “classical” inflammatory markers CRP and procalcitonin (PCT), its usefulness has been a matter of debate (Honore and Spapen 2016).

### 11.3.3 Predictive Biomarkers

Predictive biomarkers aim to objectively evaluate the outcome of a specific clinical intervention, or the differential outcomes of two or more interventions, including toxicity. Predictive and prognostic markers often get confused because many markers can be both predictive and prognostic. A marker can be an unfavorable prognostic factor but could predict favorably for response to therapy, or vice versa. Furthermore, it is possible that markers can be prognostically unfavorable, but predict favorably for one therapy and predict unfavorably for another therapy.

When searching PubMed with the joint medical subject headings “NSTI” and “predictive markers,” most, if not all, publications indexed are studies that have identified prognostic rather than predictive markers. In a recent study Afzal and coworkers (Afzal et al. 2020) performed metabolomics analysis combined with advanced bioinformatics analysis of metabolite–metabolite association networks, indicating potential of metabolites as predictive markers in NSTI (described in more detail in Sect. 11.4.5).

In conclusion, there is very little information on biomarkers that can be used to predict patient who benefits from different clinical interventions. Thus, this could potentially be an accelerating research field which would have a significant clinical impact.

### 11.3.4 Predisposing Biomarkers

The fourth category of biomarkers is used to identify individuals with a predisposition or susceptibility to develop disease. In NSTI, studies have revealed that allelic variations in human

leukocyte antigen (HLA) class II haplotypes result in striking differences in severity and outcomes of invasive GAS infections via the differential presentations of GAS superantigens (SAGs) by class II HLAs to host T-cell receptor (TCR) V $\beta$  elements (Kotb et al. 2002, 2003; Norrby-Teglund et al. 2000; Norrby-Teglund and Kotb 2000). In particular, specific HLA class II haplotypes have been found to confer strong protection from severe systemic disease, whereas others increased the risk of severe disease; patients with the DRB1\*1501/DQB1\*0602 haplotype mounted significantly reduced responses and were less likely to develop severe systemic disease.

It is likely that in NSTI a complex interaction of multiple host and microbial factors that do not lend themselves to reduction into a simple formula exist, and therefore more holistic approaches need to be implemented to identify biomarkers for diagnosis, prognosis, prediction, or predisposition/susceptibility.

---

## 11.4 Biomarker Use

### 11.4.1 Genomic Biomarkers in NSTI

To identify genomic biomarkers DNA sequencing (host and pathogen) technologies, such as whole-genome sequencing can be used. While cancer research has reached the farthest using whole-genome sequencing (Staaf et al. 2019), it is difficult to find examples where DNA sequencing has been used to identify biomarkers in NSTI. It is likely that NSTI, with its acute course, poses other challenges than cancer and in NSTI perhaps gene sequencing may be best used to identify predisposing conditions. However, the relatively high costs likely prevent whole-genome sequencing analyses from becoming routine on a large scale. Instead, “exome” sequencing (targeted sequencing of protein-coding regions) may become a cost-effective way to provide access to the entire transcribed genome of individuals. In addition, genome-wide association (GWA) studies identifying risk loci are feeding into the clinical use of DNA variants. This research has been

made possible by technological developments that have expanded the throughput and reduced the cost of single nucleotide polymorphisms (SNP)-based research and allowed SNP-based techniques to identify DNA copy number variation. Nevertheless, to identify SNPs remains a significant challenge, due to the difficulty of separating these. One of the limitations of GWA studies is that it captures only common SNPs, which by themselves may only contribute a small extent of risk to develop disease. NSTI patients infected with the same bacterial strain can develop very different manifestations and a role of immunogenetics of the host in shaping the outcome of invasive streptococcal infections can be identified (Kotb et al. 2002). However, to identify rare SNPs that contribute to develop NSTI for a small percentage of patients, more cost-effective technologies will be required. Despite its disadvantages DNA sequencing may become more important for patient care over the next 5 years as the costs are reduced and capabilities extended.

### 11.4.2 Transcriptomic Biomarkers in NSTI

Mapping genotypes to phenotypes is one of the long-standing challenges in biology and medicine, and a powerful strategy for tackling this problem is performing transcriptome analysis. Transcriptomics entails the measurement and analysis of mRNA using microarray or RNA sequencing technologies. At the beginning microarrays offered tremendous opportunities by capturing parallel information about many more mRNAs than quantitative PCR techniques (RT-PCR), which mostly focus on single mRNA analyses. However, like all molecular techniques there are limitations, such as dependence on template quality (Bustin and Nolan 2004).

Although microarrays have been highly productive research tools and used in several current prognostic and predictive tests, other emerging technologies have replaced the microarrays. Next-generation DNA sequencing technology known as “RNA-seq” allows simultaneous

analysis of all RNA molecules, including alternative splice variants, mRNAs, non-coding RNAs (ncRNAs), and miRNAs. RNA-seq has revolutionized the analysis of RNAs. In a recent study Thänert et al. took advantage of the next-generation sequencing tools to enable microbial profiling using 16S rRNA sequencing with transcriptional analysis of host and microbe using dual RNA sequencing (RNA-seq) in tissue biopsies from NSTI patients (Thänert et al. 2019). The results from this study demonstrated that, despite the similar clinical presentation of NSTIs, the pathophysiology of mono- and polymicrobial etiology differs significantly and that these differences can potentially be exploited for diagnostic purposes.

The transcriptional analysis of infected tissue indicated that the gene expression profile differed significantly between monomicrobial streptococcal infections and polymicrobial NSTI. Among the genes with significantly greater expression in polymicrobial compared to monomicrobial NSTIs were those encoding ECM components like collagen, fibronectin or lumican, as well as connective tissue growth factor, proteins. On the other hand, a set of genes encoding interferon-inducible mediators such as CXCL9, CXCL10, CXCL11, MX1, and MX2 as well as the guanylate-binding GTP1 and GTP2 were most prominently higher expressed in monomicrobial NSTIs, in particular those caused by *S. pyogenes*.

Because of the many diverse cell types in our body each of which express a unique transcriptome, conventional bulk population sequencing can provide only the average expression signal for an ensemble of cells. More recently there has been an explosion of interest in obtaining high-resolution views of single-cell transcriptomics (scRNA-seq) heterogeneity. Assessing the differences in gene expression between individual cells has the potential to identify rare populations that cannot be detected from an analysis of pooled cells. Due to issues with cost and analysis challenges, the implementation of scRNA-seq as a tool for biomarker identification in NSTI may still be some years away and is unlikely to give the complete picture.

### 11.4.3 Epigenomic Biomarkers in NSTI

Epigenetics can be defined as the field of inheritable changes in gene expression that are not caused by alterations in DNA sequences. Epigenetic mechanisms include DNA methylation, histone modifications, and non-coding RNAs, and epigenetic alterations have an increasingly clear role in modulating inflammatory and other immunological processes. Identification of epigenetic alterations may provide important prognostic, predictive markers. This is an exciting field that is likely to grow in clinical importance.

There are no studies on epigenetics in NSTI, while the often-associated condition sepsis has been subjected to epigenetic analyses (Cross et al. 2019). To date, mostly the epigenetic modification associated with various stages of sepsis has been assessed. A focus has been to identify mechanisms involved in endothelial dysfunction during the hyperinflammatory response and those underpinning aspects of immunosuppression.

### 11.4.4 Proteomic Biomarkers in NSTI

Large scale analysis of specific proteomes, also known as proteomics, defines protein diversity and understands its biological consequences. Analysis of proteins as inflammatory markers in diseases could provide important information about disease severity and guide decision making. In true proteomics analysis all proteins from a sample of interest are usually extracted and digested with one or several proteases (typically trypsin alone or in combination with Lys-C (Wiśniewski and Mann 2012) to generate a defined set of peptides. The peptides obtained are subsequently analyzed by liquid chromatography coupled to mass spectrometry (LC-MS).

The two most common approaches are either designed to achieve a deep coverage of the proteome (shotgun MS) or to collect as much quantitative information as possible for a defined set of proteins/peptides (targeted MS) (Picotti and Aebersold 2012). Protein content information can also be obtained through targeted approaches,



including multiplex analyses made to identify a set of predetermined proteins. Since the test results are needed within hours in NSTI, approaches will most realistically include analyses of samples, such as plasma, that can be obtained quickly. However, rapid assessment based on proteomics and/or multiplex analyses of NSTI patients to identify protein-based biomarkers is challenging, as these are time-consuming analyses. Nevertheless, multiplex approaches have been used to analyze NSTI plasma samples and has provided important insight to potential diagnostic biomarkers (Thänert et al. 2019; Hansen et al. 2017a).

#### 11.4.5 Metabolomic Biomarkers in NSTI

Metabolomics takes a special position among the omics disciplines in the system top-down approach as the metabolome of biological processes, carrying imprints of genetic, epigenetic, and environmental factors, provides the link between the genotype and phenotype (Karakitsou et al. 2019; Rosato et al. 2018; Dunn and Ellis 2005; Kell 2004). Since metabolites represent the endpoint of many molecular pathways, metabolomics which quantifies the metabolite content of cells or tissues is in addition to have diagnostic and prognostic value also potentially useful for predicting treatment response. Metabolomics has been identified as a novel tool to discover targets for sepsis diagnosis and prognosis, as well as to gain insight into pathogenic disease mechanisms (Eckerle et al. 2017).

Recently, we used untargeted metabolomics analyses of plasma from NSTI patients and healthy controls to identify the metabolic signatures and connectivity patterns among metabolites associated with NSTI. Out of 97 metabolites detected, the abundance of 33 was significantly altered in NSTI patients. Analysis of metabolite–metabolite association networks identified 20 metabolites differently connected between NSTI and controls. Testing of a set of differently connected metabolites (ornithine, ribose, urea, and glucuronic acid) in *in vitro*

biofilm formation assays revealed metabolite-specific effects on both bacterial growth and biofilm formation. A biofilm can be considered as a potential complicating microbiological feature of NSTI and, consequently, emphasizes reconsideration of antibiotic treatment protocols. This study identifies for the first time an NSTI-specific metabolic signature with implications for optimized diagnostics and therapies. Thus, the finding of metabolites in NSTI associated with biofilm formation emphasizes the potential of metabolites as predictive markers in NSTI.

#### 11.4.6 Microbiomic Biomarkers in NSTI

Analyzing the microbiome at the system biology level for comparison in diseases to that of healthy subjects has become increasingly popular. Rapid progress in the development of next-generation sequencing (NGS) technologies in recent years has provided many valuable insights into complex biological systems, including diverse microbial communities.

As far as NSTI is concerned, this field of research is in its infancy, but it may prove helpful in providing biomarkers which will give insights into the pathological process behind disease evolution and progression by determining specific etiological factors.

#### 11.4.7 From One-Category-Analyses Towards Systems Biological Approaches

The classical approach to biomarker discovery usually entails the comparison of molecules or group of molecules among two or more groups, usually in a case/control scenario. In NSTI typical comparison is monomicrobial vs polymicrobial infection or patient outcomes (septic shock/no-septic shock (Hansen et al. 2018), amputation/no-amputation, or mortality (Hansen et al. 2018)) or some clinical score (low/high LRINEC) (Hansen et al. 2017a). Statistical approaches are usually simple: candidate biomarkers are mostly

tested one at a time in a univariate fashion using procedure like ANOVA or linear mixed models, logistic regression, or Kaplan–Meier curves when association with survival is of interest or by correlating biomarker concentration with some clinical outcome.

Most new biomarkers proposed in the literature never reach the clinic, often because of a lack of reproducibility. In a meta-analysis of highly cited articles announcing new biomarker candidates for a variety of diseases, it was shown that follow-up experiments with greater statistical power generally fail to reproduce the same effect size as the original studies (Ioannidis and Panagiotou 2011; Feala et al. 2013). Moreover, in addition to the need for validation of such candidates, integration of the information provided by each biomarker is also needed for a comprehensive representation of the disease process. Here we argue that biomarker discovery would be greatly enhanced by applying systems biology principles such as the multi-scale integration of information and the analysis of dynamic patterns with the help of computational tools.

---

## 11.5 Systems Biology

Systems biology is an interdisciplinary field in which experimental and computational methods concur with mathematical models to the analysis and the understanding of biological phenomena. In systems biology different types of molecular knowledge are integrated and exploited thanks to the synergistic use and combination of statistical and mathematical model together with experimental data (Bruggeman and Westerhoff 2007).

Systems biology embraces a system prospective as opposed to the reductionistic vision which is common in the mainstream biology: it aims to tackle biological problems by considering the interaction among the different parts of the systems rather than subdividing it into independent sub-problems. In biological systems the interactions between components dominate the component themselves in shaping the system-wide behavior and for this reason the reductionist approach is less effective to understand the

biomolecular mechanisms underlying complex biological processes (Ahn et al. 2006)

Systems biology relies on the use of a mathematical model to integrate and understand data that are acquired using a vast array of experimental approaches ranging from molecular techniques to comprehensive measurement often encompassing multiple layers of the omics cascade such as transcriptomics, proteomics, and metabolomics. Different types of models can be used depending on the data available, the research questions, and whether a deductive or an inductive approach is followed.

The deductive approach, usually referred to as *Top-down* approach (Oltvai and Barabási 2002), aims to gain insights into a given phenomenon starting from the system-wide data acquired using omics technologies from which information is extracted, usually using statistical and machine learning methods in combination with network inference and network analysis (Rosato et al. 2018; Ideker and Krogan 2012). The ultimate goal is to describe the interactions among the molecular constituents of the system (genes, proteins, metabolites), possibly across different conditions (Rosato et al. 2018, Ideker and Krogan 2012) to understand how these parts interact to shape the system-wide behavior of the system.

On the other side, the inductive approach, referred to as *Bottom-up* approach (Oltvai and Barabási 2002), starts from a detailed molecular and biochemical knowledge of certain biological mechanisms and aims to create mathematical models that can reproduce experimental data. The overarching goal is to predict the system-wide behavior by building a complete genome-scale model to provide an integrative view of the biological interactions occurring inside living systems (Shahzad and Loor 2012). The iterative cycle of systems biology is illustrated in Fig. 11.1 in relationship to the pipeline for biomarker discovery.

## 11.6 Systems Biology Approaches to Biomarker Discovery

Determining whether a given biomarker is causal or reactive to NSTI (or, in general, to any disease) is likely to provide information on the underlying pathogenic process and ultimately impact its general applicability (Villoslada and Baranzini 2012). However, standard approaches that are based on just univariate analysis can fail, by their own nature, to account for the interrelatedness existing among genes, proteins, and metabolites that behave in an orchestrated way for what concerns regulation, transcription, and translation.

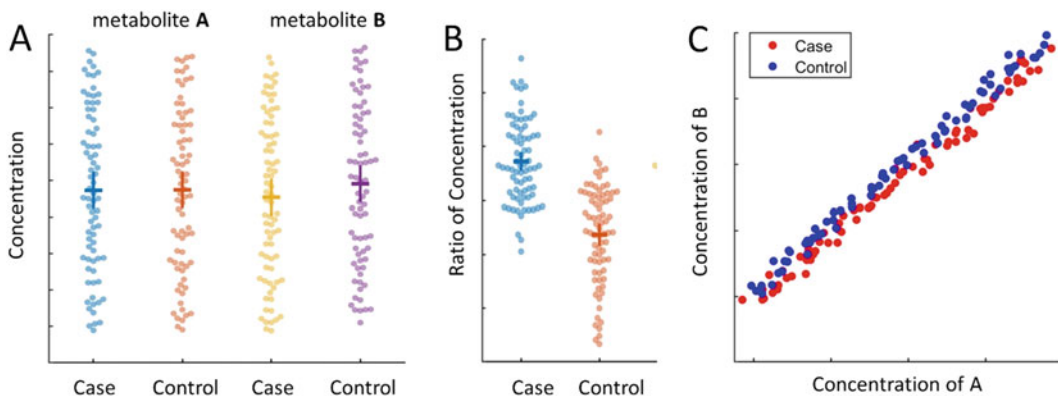
For instance, Edwards and co-workers (Edwards et al. 2018) investigated *S. pyogenes* infections including NSTI among others, using a proteomic approach and identified several human proteins associated with NSTI. Yet, data was analyzed in a univariate fashion, missing the potential of highlighting association patterns among proteins that could have been uniquely associated with one or more type of infections.

Systems biology approaches rely on the concept of mutual interaction among the constituents of a biological system and thus poses a great emphasis on measuring, describing, analyzing, and interpreting the relationships among molecular features (genes, proteins, metabolites...) rather than considering them one at a time. In this light, top-down systems biology approaches can be considered multivariate in nature, that is, the analysis is performed taking into consideration the relationships among the variables and not only their mean levels (i.e. concentration and abundances) (Saccenti et al. 2014a) as in the standard analysis for single molecule biomarkers. Relationships among genes, proteins, and metabolites are quantified using association measures which describe the similarity of the concentration and abundance profiles of these molecular features measured on a set of biological replicate samples. Correlations (either the Pearson's or Spearman's indexes) are commonly used and biological and biochemical information can be derived from both the strength and the sign

of correlations; for instance, if analyzing metabolites, a strong positive correlation can indicate an equilibrium condition or enzyme dominance, while strong negative correlation can indicate the presence of a conserved moiety (Camacho et al. 2005).

It is fundamental to understand that the associations among variables can carry information also when concentrations (and abundances) do not carry relevant information, for instance, to discriminate between the case and control groups. This situation, so common in the biomarker discovery setting, is better illustrated with a simulated example. With reference to Fig. 11.2a, consider two metabolites (or protein or any other biomolecular quantity) A and B measured from case patients and controls: under this simulation there is no difference between the concentration levels of A and B in the two groups. However, if the case and control group are compared taking ratio of the concentrations of A and B a clear difference between case and controls is evident, as shown in Fig. 11.2b (see figure caption for more details) and it is clear that the information discriminating between cases and controls relies on both metabolites and their mutual relationship

This phenomenon is a consequence of metabolites A and B being correlated as shown in Fig. 11.2c: from biological point of view what is observed here is that given a certain concentration of metabolite A, the concentration of metabolite B changes with class (Saccenti et al. 2014a). This indicates that when the relation or interaction between pairs of metabolites or other biomolecular features is thought to be important and to carry information about the problem studied, the ratios of metabolite concentration (or abundances) could be considered instead of single metabolite concentrations. Examples of this exist in the clinical setting where the deviation from the normal range of the ratios of metabolites or protein is used as diagnostic biomarkers (Petersen et al. 2012): ornithine- $\delta$ -aminotransferase deficiency in young children is diagnosed using the proline to citrulline ratio (De Sain-Van Der Velden et al. 2012), while the ratio between blood phenylalanine and tyrosine concentrations is used to

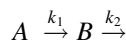


**Fig. 11.2** Idealized experiment aiming to find biomarkers discrimination between the case and control group based on the measurement and analysis of two metabolites A and B measure on  $n_1 = 75$  cases and  $n_2 = 75$  controls. (a) Comparison of the concentrations of metabolites A and B tested with 1-way ANOVA: there is no statically significant difference at the  $\alpha = 0.05$  level ( $P$ -value 0.32 and

0.98, respectively, for metabolite A and B. (b) Comparison of the ratios of the concentrations of metabolites A and B: a discrimination between the case and control groups is evident (1-way ANOVA  $P$ -value =  $2.7 \times 10^{-18}$ ). (c) Scatter plot of the concentration profiles of metabolites A and B

identify heterozygous carriers of phenylketonuria risk alleles (Hsia 1958).

Metabolite and protein ratios and variation thereof have often direct biological relevance and can provide mechanistic insights. In fact, if two metabolites are connected within (or by) a biochemical pathway, metabolite ratios approximate (under idealized steady state assumptions) the reaction rates (Petersen et al. 2012). For instance, considering the elementary biochemical reaction where metabolite A is converted to B and then degraded



described by the differential equations

$$\begin{cases} \frac{d[A]}{dt} = -k_1[A] \\ \frac{d[B]}{dt} = +k_1[A] - k_2[B] \end{cases}$$

where  $[A]$  and  $[B]$  are the concentrations of A and B at time  $t$ . At steady state the ratio of the concentration of A and B gives an approximation of the ratios of the rate constants for the two reactions

$$\frac{[A]}{[B]} = \frac{k_2}{k_1}$$

and can be used as a proxy to characterize the system. Since reaction constants are tightly constrained under normal physiological conditions, increased (or decreased) metabolite ratios can be an indication of alteration in the activity of an enzyme that catalyzes the reaction or a change in flux distribution. For instance, variation of the ratios between sphingolipids that differ by two carbon moieties has been linked to a modified beta-oxidation, while variation of the ratios between different classes of phospholipids has been proposed to describe modified activity of enzymes in the phospholipid pathways (Altmaier et al. 2008).

## 11.7 Networks and Network Analysis

When many metabolites or other biochemical and biomolecular features are measured, their mutual relationships can be conveniently represented using networks. A (biological) network is a graphical representation of objects (called nodes) and their relationships described by links

or edges connecting different nodes. The exact meaning of the nodes and edges, i.e. of the kind of relationships they represent, depends on the context. In gene regulatory networks the nodes represent genes and the edge represents the existence of regulatory mechanism where the protein produced by a given gene regulates the expression of the target gene (Emmert-Streib et al. 2014). In networks describing protein–protein interaction the edge represents the existence of a physical interaction between two proteins (Jordán et al. 2012).

In the context of biomarker discovery, the most relevant networks are probably the so-called metabolite–metabolite association networks, or correlation networks, where the nodes are metabolites and the edges represent the existence of associations between two metabolites, quantified through the correlation of their concentration profiles measured on replicated samples.

The simplest approach to build metabolite–metabolite association networks is first taking the correlation  $r_{ij}$  among pairs of metabolites  $i$  and  $j$  and then imposing a threshold on either the magnitude of the correlation or on the statistical significant (or both), setting to 0 the correlation that do not satisfy such conditions. For instance

$$r_{ij} = \begin{cases} r_{ij} & \text{if } |r_{ij}| > \theta \text{ and } P\text{-val} < \alpha \\ 0 & \text{otherwise} \end{cases} \quad (11.1)$$

where  $\theta$  and  $\alpha$  (statistical significance) are parameters to be determined; a commonly accepted value for  $\theta$  is 0.6 which is usually taken as a threshold to discriminate between low and medium-high correlation (Camacho et al. 2005). Other methods, like the Probabilistic Context Likelihood of Relatedness for Correlations (PCLRC) (Saccenti et al. 2014b), dispose of the necessity of a pre-defined threshold on the correlation but are context dependent, and the significance of the correlations is established with respect to the correlation background observed in the data.

If the value of the correlation (or any other association measure) is retained, the network is said to be weighted as every edge connecting two nodes describes the weight of the association. In other cases, only the presence/absence of an association is retained, and in this case the network is said to be unweighted

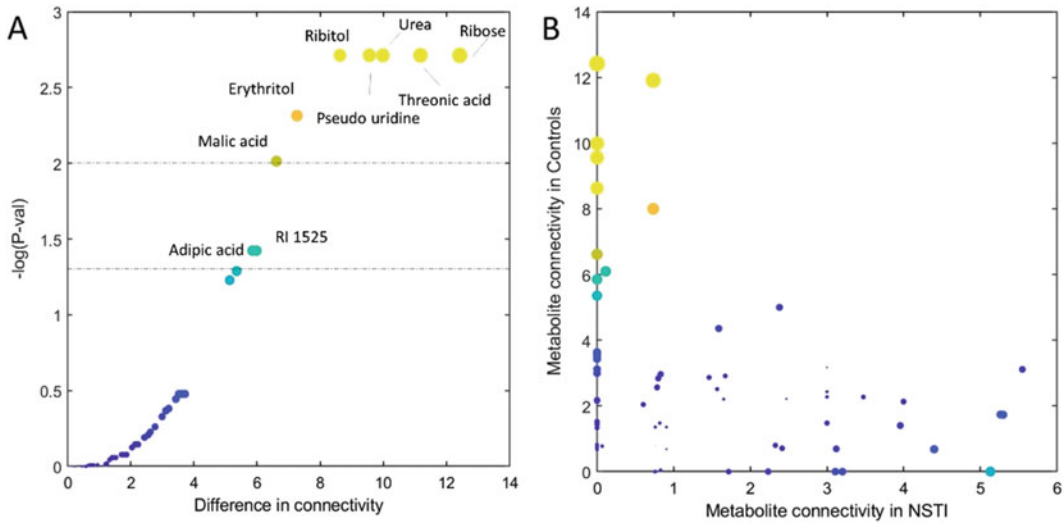
$$r_{ij} = \begin{cases} 1 & \text{if } |r_{ij}| > \theta \text{ and } P\text{-val} < \alpha \\ 0 & \text{otherwise} \end{cases} \quad (11.2)$$

The nodes in a network can be characterized using functions that can be derived from the patterns of its association: a common measure is the node degree or connectivity. For unweighted networks this is simply the number of its connection. For weighted networks a weighted connectivity is defined as the sum of the absolute values of the edges:

$$\chi_i = \sum_{i>j} |r_{ij}| \quad (11.3)$$

In the framework of biomarker discovery, the patterns of correlations between metabolites and their connectivity can be exploited by comparing them across conditions, for instance, cases and controls, to identify associations that are disrupted or altered by pathophysiological conditions. This approach is called differential network analysis and has been applied in NSTI research (Afzal et al. 2020) as well in other fields like cardiovascular disease (Vignoli et al. 2020; Saccenti et al. 2014b) and aging (Vignoli et al. 2018).

Afzal et al. (2020) combined classical univariate and multivariate approaches together with differential analysis of metabolite–metabolite correlation networks which were built from metabolite concentrations of blood plasma metabolites measured using GC-MS from NSTI patients ( $n = 34$ ) and healthy (noninfected) controls ( $n = 24$ ) enrolled in the INFECT project (Madsen et al. 2019). In this study Spearman’s rank correlation was used as an index of metabolite association and implemented a conservative selection procedure involving resampling to obtain robust



**Fig. 11.3** (a) Absolute differences in metabolite connectivity in networks from NSTI and healthy control subject (see Eqs. (11.3) and (11.4)) against its statistical significance (Benjamini–Hochberg corrected  $-\log_{10}(P\text{-value})$ ). The horizontal lines represent the significance thresholds at  $\alpha = 0.05$  and  $0.01$  level. (b) Scatter plot of the

metabolite connectivity observed in the metabolite–metabolite correlation networks built from NSTI and noninfected surgical controls. The colors are proportional to  $-\log_{10}(P\text{-value})$  and the marker size is proportional to the difference in connectivity

estimates of the correlation because of the small sample size available. Metabolite association networks were constructed for both NSTI and healthy and for each of the 97 measured metabolites they calculated the differential connectivity  $\Delta\chi_i$  defined as

$$\Delta\chi_i = \chi_i^{\text{NSTI}} - \chi_i^{\text{Control}} \quad (11.4)$$

where  $\chi_i^{\text{NSTI}}$  and  $\chi_i^{\text{Control}}$  are the connectivity for the metabolite  $i$  in the association network specific for NSTI and healthy control, respectively. The significance of the observed difference in connectivity was assessed using a permutation test. Figure 11.3a shows the results of differential network analysis and Fig. 11.3b shows the metabolite connectivity in the NSTI specific network against connectivity in the uninfected surgery controls. The authors noted that several metabolites (urea, ribose, ribitol, pseudo uridine, malic acid) that were not connected in the control network were highly connected in the

NSTI network, becoming hub node. In network biology a hub is a node that is densely connected, i.e. it is associated with many other nodes. Because of this, hub nodes are assumed to play crucial biological roles.

The concept of hub was first introduced within the analysis of yeast protein–protein interaction networks and it was shown that highly connected nodes are more likely to be essential for survival (Jeong et al. 2001; Carter et al. 2004). Analysis of hub nodes has been extended to metabolite–metabolite association networks: comparing low and high cardiovascular risk patients. Saccenti et al. (2014b) identified conserved hub metabolite, and differentially conserved hubs and linked them to alteration of mitochondrial activity.

Given this, Afzal et al. speculated that metabolites showing differential connectivity could play a role in disease pathogenesis either by affecting the host response or the bacterial infection. A set of significantly altered metabolites was selected and tested in vitro to

investigate potential influence on NSTI group A streptococcal strain growth and biofilm formation.

The selected metabolites (ornithine, ribose, urea, and glucuronic acid) were supplemented to chemically defined media and revealed metabolite-specific effects on both bacterial growth and biofilm formation.

---

## 11.8 Multivariate Biomarker and Omics Signatures

A generalized class of biomarkers are the so-called multivariate biomarkers which are a generalization of the bivariate biomarker described in Sect. 11.6. This kind of biomarkers are sometimes denoted with the term *signatures*, especially in the context on *omics* research where gene, protein and metabolite expression, abundances and/or concentration levels are measured in a comprehensive fashion on biological samples using high-throughput analytical techniques like RNA sequencing, mass spectrometry, or nuclear magnetic resonance spectrometry. These experimental techniques are multivariate in nature since hundreds to thousands of molecular features are measured simultaneously and such molecular signatures are exploited through multivariate and machine learning approaches.

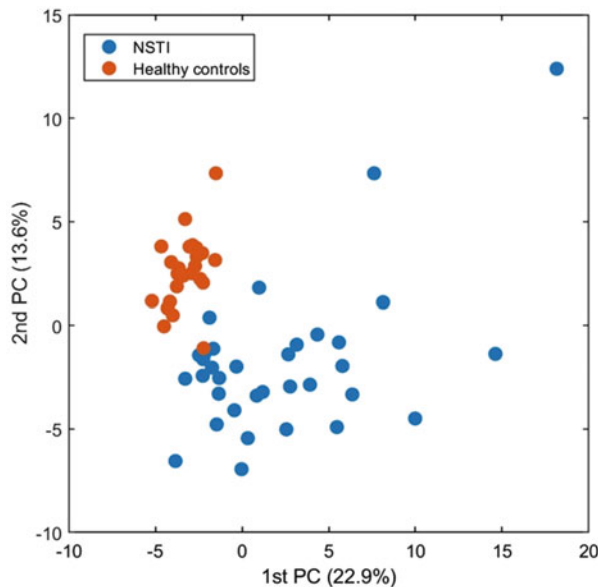
Sung et al. (2012) defined operatively molecular signature as *a set of biomolecular features together with a pre-defined computational procedure that applies those features to predict a phenotype of clinical interest on a previously unseen patient sample*, stressing the computational/statistical nature of such signatures.

Signatures are defined using statistical approaches in combination with machine learning techniques. Principal component analysis (PCA) (Hotelling 1931; Pearson 1901; Jolliffe 2002) is a common starting point when looking for biomarker signatures and it is used to explore data and to highlight if such a signature may exist. PCA is data reduction technique that allows for visualization and exploration of high-dimensional data by reducing its dimensionality. This is

accomplished by taking linear combinations of the original variables that explain as much as data variability as possible which allows to visualize the data in a lower dimensional space. Figure 11.4 shows the results of a PCA analysis on the data from Afzal and colleagues (Afzal et al. 2020). A clear separation between the metabolite profiles of NSTI patient and healthy controls is evident, indicating the existence of a metabolomics (in this case) signature as shown in the PCA score plot in Fig. 11.4. Figure 11.5 shows the relative importance (i.e. the PCA loadings) corresponding to the PCA score plot. The loadings are measure of the contribution of each variable to explain the variability observed in the data, in this case the separation between NSTI and controls. As it can be seen all variables (only the first 40 are shown for simplicity) contribute to the model, showing the multivariate nature of this metabolomics signature differentiating NSTI subjects from the controls. It is interesting to note that among the most important metabolites there are urea, ribose, ribitol, pseudo uridine penta, malic acid, the same metabolites that showed significant differential connectivity in the network analysis (see Fig. 11.3). This is because PCA, like all multivariate approaches, takes into account the correlation among variables which are also reflected in the network structure underlying the data, making network analysis and multivariate statistics two complementary approaches.

Omics signatures consisting of many biomolecular features have been defined and routinely applied in the clinical setting. For instance, the MammaPrint test (Van 'T Veer et al. 2002) used to determine the most appropriate chemotherapy treatment in lymph node negative breast cancer patients with either positive or negative ER, consists of 70 gene expression signatures which were found to correlate with clinical phenotype (metastatic vs. non-metastatic) and validated on independent patient cohorts (Van de Vijver et al. 2002). Given the particular nature of NSTI, which is a fast spreading infection requiring extremely fast diagnosis and treatment, this type of high-dimensional genomics signatures may not of immediate applicability in the clinical setting

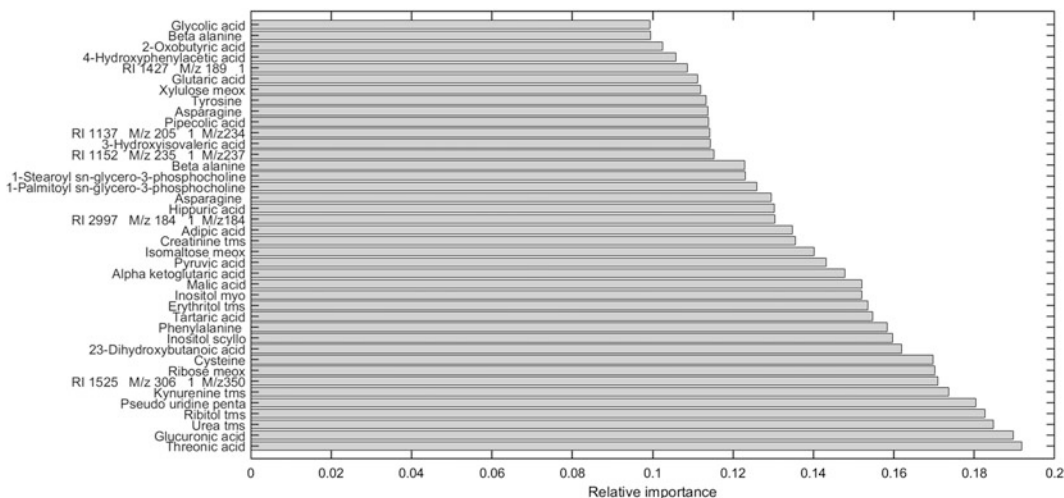
**Fig. 11.4** Score plot of principal component analysis of concentration profiles of 97 metabolites measured on using GC-MS from NSTI patients ( $n = 34$ ) and healthy (surgery, noninfected) controls ( $n = 24$ ) enrolled in the INFECT project (Madsen et al. 2019). Data is from (Afzal et al. 2020). A clear separation between NSTI patient and controls is evident, indicating the existence of a multivariate metabolomics signature specific to NSTI. See Fig. YY for the corresponding loading plot



although some applications are emerging (See also Chap. 13, Sect. 11.3.2).

However, omics signature can be explored to single out key players and provide mechanistic information about disease onset and progression and can generate hypotheses that can be further

test in vitro and or in vivo in animal models, closing the iterative cycle typical of systems biology approaches. In this respect the main limitation of PCA based signatures is that (1) all variables contribute to the model and (2) PCA base signatures are exploratory in nature and do



**Fig. 11.5** Loading plot of principal component analysis of the concentration profiles of 97 metabolites measured on using GC-MS from NSTI patients ( $n = 34$ ) and healthy (surgery, noninfected) controls ( $n = 24$ ) enrolled in the INFECT project (Madsen et al. 2019). Data is from (Afzal

et al. 2020). Only the first 40, most contributing metabolites are shown. Urea, ribose, ribitol, pseudo uridine penta, and malic acid are also found to be relevant from network analysis, See Fig. 11.4 for the corresponding score plot



not allow, per se, the prediction of the clinical phenotype on previously unseen patient samples.

Signatures that are easier to interpret can be obtained using advanced explorative techniques like sparse approaches (Camacho et al. 2017; Saccenti et al. 2018; Camacho et al. 2020). The existence of a predictive power of such signatures can be investigated using chemometrics or machine learning approaches like partial least squares discriminant analysis (PLS-DA) (Wold and Eriksson 2001) or some of its extensions (Lê Cao et al. 2008; Bylesjö et al. 2006; Camacho and Saccenti 2018), Random Forests (Breiman 2001), Support vector machines (Cortes and Vapnik 1995), or other machine learning tools.

When properly implemented, these methods can allow unbiased prediction of the clinical phenotype on previously unseen patient samples: this is accomplished by implementing cross-validation, which is a computational strategy that implies removing a set of samples from the data set, building a predictive model, and then testing in an unbiased way the ability of the model to predict the left out samples, mimicking, in this way, validation on an external cohort of samples. This procedure can be particularly relevant in NSTI where validation cohorts are difficult, to obtain. Random Forest was used to define a multivariate biomarker consisting of 11 blood analytes attaining specificity and sensitivity around 90% for the early detection of ovarian cancer (Bertenshaw et al. 2008), while Support vector machine has been used to derive a panel of 5 markers able to attaining a prediction sensitivity of 95.3% and a specificity of 99.4% (Mor et al. 2005).

Predictive *omics* signatures have great potential in NSTI. As an example, we show here a Random Forest classification model fit on the data from (Afzal et al. 2020) with default parameters and implemented by splitting the data into a validation set (20% of the data) and a training set (80%) and repeating the cross-validation procedure 100 times to take into account sampling variability results. On the training set the accuracy for the prediction of NSTI and control samples is 0.981 (0.978, 0.982 95%

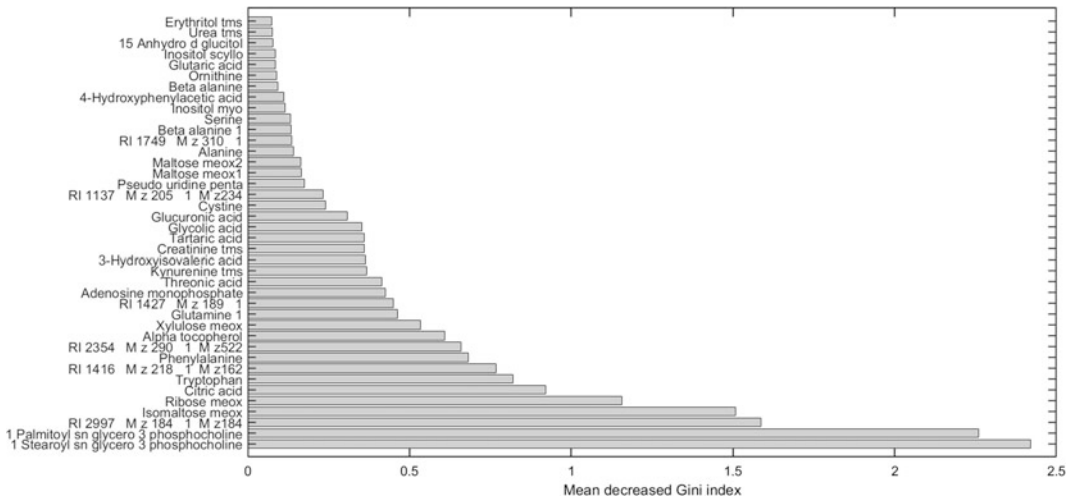
CI), while on the validation set one get average accuracy equal to 0.980 (0.975–0.995, 95% CI), average sensitivity 0.954 (0.938–0.969, 95% CI), and an average specificity of 1. The classification results on the validation set are similar to those on the training set and this indicates that, in principle, these results may be generalizable to samples acquired on external validation cohorts.

Many classification tools also produce measures of importance of the predictors used in the model and this allows to single out which variables are most contributing to the discrimination among the groups to gain understanding of the underlying biological processes or identify potential biomarkers. Figure 11.6 shows the variable importance plot of the Random Forest discriminant model previously described to discriminate between NSTI and healthy controls. It is interesting to note that the group of variables selected as important is somehow different from the one from the PCA analysis (see Fig. 11.5): only ribose is present in the top five most important variables, while urea, which is important in the PCA model, is not important in the Random Forest model. This should not be surprising, since PCA and Random Forest have different aims and exploit different characteristics of the data: PCA is an unsupervised method, while Random Forest is a supervised method that uses information about sample class (i.e. whether they belong to NSTI patient or to controls) to build a model able to predict to which class an unknown sample may belong.

---

## 11.9 Future Perspective

Technology advances and critical analysis of molecular pathways have opened new horizons for management of diseases, exploring therapeutic solutions to each individual patient beyond the one-size fits all practice. As the amount of information available grows, there comes a stage when there is too much information to manually integrate reliably. This leads to the need for formal mathematical or computational models of disease that incorporate all this information. For this reason, clinical, microbiological, and experimental



**Fig. 11.6** Variable importance plot of a Random Forest discriminant model fit to the concentration profiles of 97 metabolites measured on using GC-MS from NSTI patients ( $n = 34$ ) and healthy (surgery, noninfected) controls ( $n = 24$ ) enrolled in the INFECT project (Madsen

et al. 2019). Data is from Afzal et al. (2020). Only the first 40, most contributing metabolites are shown. The model is used to investigate the predictive capability of the metabolomics signature discriminating NSTI patient from controls

data can be integrated using various types of mathematical algorithms to provide decision support tools for clinicians. This can pave the way for application of personalized medicine in infectious diseases, such as NSTI, with the aims to achieve the right diagnosis and right treatment for the right patient at the right time at the right cost.

Multivariate statistics, machine learning, and reverse engineering approaches will allow pinpointing and mapping key nodes that can reveal possible biomarker sets from heterogeneous (meta-)data both from the tested pathogens and the hosts (including the various “omics,” pathotyping, patient stratification data). These top-down analyses will be iteratively combined with bottom-up modelling of specific sub-networks/biomarkers. By thoroughly applying an iterative “dry-wet” cycle centered around the cross-linking of bottom-up, hypothesis-driven modelling of selected pathways and/or signatures, with top-down, unbiased systems biology approaches biomarkers that may contribute to disease onset and outcome can be identified. In particular, biomarker discovery in NSTI will benefit from the integration of heterogeneous data, such as patient data and metadata, data from the tested pathogens and the murine models

and the wealth of omics data that can be collected from them.

Even if the vast heterogeneity and/or quality of the data will not allow detailed mechanistic modelling of specific pathways, these top-down, unbiased approaches will at least enable to identify basic structures in the topology of the various networks underlying the disease. This in itself is a substantial contribution to the state of the art in biomarker identification. Finally, it will be important to validate identified signatures/biomarkers for patient classification through the use of large patient cohorts. Indeed, validation remains one of the major bottlenecks for biomarker discovery. To become a clinically approved test, a potential biomarker should be confirmed and validated using large cohorts and should be reproducible, specific, and sensitive (Drucker and Krapfenbauer 2013). Although not uncommon, NSTI is still a rare condition whose characteristics (difficult diagnosis, fast spread, high mortality, among others) make the creation of validation cohorts difficult. However, the INFECT cohort and the methodologies and strategies for data collection and analysis developed within the INFECT project (Madsen et al. 2018, 2019; Afzal et al. 2020) are a first and

promising step towards the definition of clinically relevant biomarker in NSTI.

**Acknowledgement** Financial support: The work was supported by the European Union Seventh Framework Programme: (FP7/2007–2013) under the grant agreement 305340 (INFECT project); the Swedish Governmental Agency for Innovation Systems (VINNOVA) under the frame of NordForsk (Project no. 90456, PerAID), and the Swedish Research Council and The Netherlands Organization for Health Research and Development (ZonMv) under the frame of ERA PerMed (Project 2018-151, PerMIT).

## References

- Alfzal M, Saccenti E, Madsen MB, Hansen MB, Hyldegaard O, Skrede S, Martins Dos Santos VAP, Norby-Teglund A, Svensson M (2020) Integrated univariate, multivariate, and correlation-based network analyses reveal metabolite-specific effects on bacterial growth and biofilm formation in necrotizing soft tissue infections. *J Proteome Res* 19:688–698. <https://doi.org/10.1021/acs.jproteome.9b00565>
- Ahn AC, Tewari M, Poon C-S, Phillips RS (2006) The clinical applications of a systems approach. *PLoS Med* 3:e209
- Altmajer E, Ramsay SL, Graber A, Mewes H-W, Weinberger KM, Suhre K (2008) Bioinformatics analysis of targeted metabolomics—uncovering old and new tales of diabetic mice under medication. *Endocrinology* 149:3478–3489
- Anvar MS, Minuchehr Z, Shahlaei M, Kheitan S (2018) Gastric cancer biomarkers; a systems biology approach. *Biochem Biophys Rep* 13:141–146
- Bertenshaw GP, Yip P, Seshaiiah P, Zhao J, Chen T-H, Wiggins WS, Mapes JP, Mansfield BC (2008) Multianalyte profiling of serum antigens and autoimmune and infectious disease molecules to identify biomarkers dysregulated in epithelial ovarian cancer. *Cancer Epidemiol Biomarkers Prev* 17:2872–2881
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32
- Bruggeman FJ, Westerhoff HV (2007) The nature of systems biology. *Trends Microbiol* 15:45–50
- Bustin SA, Nolan T (2004) Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *Jf Biomol Tech* 15:155
- Bylesjö M, Rantalainen M, Cloarec O, Nicholson JK, Holmes E, Trygg J (2006) OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J Chemom* 20:341–351
- Camacho J, Saccenti E (2018) Group-wise partial least square regression. *J Chemometr* 32:e2964
- Camacho D, de la Fuente A, Mendes P (2005) The origin of correlations in metabolomics data. *Metabolomics* 1:53–63
- Camacho J, Rodríguez-GÓMEZ RA, Saccenti E (2017) Group-wise principal component analysis for exploratory data analysis. *J Comput Graph Stat* 26:501–512
- Camacho J, Smilde AK, Saccenti E, Westerhuis JA (2020) All sparse PCA models are wrong, but some are useful. Part I: computation of scores, residuals and explained variance. *Chemom Intel Lab Syst* 196:103907
- Carter S, Brechbuhler C, Griffin M, Bond A (2004) Gene co-expression network topology provides a framework for molecular characterization of cellular state. *Bioinformatics* 20:2242–2250
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20:273–297
- Cross D, Drury RE, Hill JL, Pollard AJ (2019) Epigenetics in sepsis: understanding its role in endothelial dysfunction, immunosuppression and potential therapeutics. *Front Immunol* 10:1363
- De Sain-Van Der Velden MGM, Rinaldo P, Elvers B, Henderson M, Walter JH, Prinsen BHCMT, Verhoeven-Duif NM, DE Koning TJ, van Hasselt P (2012) The Proline/Citrulline ratio as a biomarker for OAT deficiency in early infancy. *JIMD Rep* 6:95–99
- Del Campo M, Jongbloed W, Twaalfhoven HA, Veerhuis R, Blankenstein MA, Teunissen CE (2015) Facilitating the validation of novel protein biomarkers for dementia: an optimal workflow for the development of sandwich immunoassays. *Front Neurol* 6:202
- Drucker E, Krapfenbauer K (2013) Pitfalls and limitations in translation from biomarker discovery to clinical utility in predictive and personalised medicine. *EPMA J* 4:7
- Dunn WB, Ellis DI (2005) Metabolomics: current analytical platforms and methodologies. *Trends Anal Chem* 24:285–294
- Eckerle M, Ambroggio L, Puskarich MA, Winston B, Jones AE, Standiford TJ, Stringer KA (2017) Metabolomics as a driver in advancing precision medicine in sepsis. *Pharmacotherapy* 37:1023–1032
- Edwards RJ, Pyzio M, Gierula M, Turner CE, Abdul-Salam VB, Sriskandan S (2018) Proteomic analysis at the sites of clinical infection with invasive *Streptococcus pyogenes*. *Sci Rep* 8:5950
- Emmert-Streib F, Dehmer M, Haibe-Kains B (2014) Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. *Front Cell Dev Biol* 2:38
- Feala JD, Abdulhameed MDM, Yu C, Dutta B, Yu X, Schmid K, Dave J, Tortella F, Reifman J (2013) Systems biology approaches for discovering biomarkers for traumatic brain injury. *J Neurotrauma* 30:1101–1116
- Goossens N, Nakagawa S, Sun X, Hoshida Y (2015) Cancer biomarker discovery and validation. *Transl Cancer Res* 4(3):256–269
- Hajar R (2016) Evolution of myocardial infarction and its biomarkers: a historical perspective. *Heart Views* 17:167
- Hansen MB, Rasmussen LS, Garred P, Bidstrup D, Madsen MB, Hyldegaard O (2016a) Pentraxin-3 as a

- marker of disease severity and risk of death in patients with necrotizing soft tissue infections: a nationwide, prospective, observational study. *Crit Care* 20:40
- Hansen MB, Rasmussen LS, Pilely K, Hellemann D, Hein E, Madsen MB, Hyldegaard O, Garred P (2016b) The lectin complement pathway in patients with necrotizing soft tissue infection. *J Innate Immun* 8:507–516
- Hansen MB, Rasmussen LS, Svensson M, Chakrakodi B, Bruun T, Madsen MB, Perner A, Garred P, Hyldegaard O, Norrby-Teglund A, INFECT Study Group (2017a) Association between cytokine response, the LRINEC score and outcome in patients with necrotising soft tissue infection: a multicentre, prospective study. *Sci Rep* 7:42179–42179
- Hansen MB, Rasmussen LS, Svensson M, Chakrakodi B, Bruun T, Madsen MB, Perner A, Garred P, Hyldegaard O, Norrby-Teglund A, INFECT Study GROUP, Nekludov M, Arnell P, Rosén A, Oscarsson N, Karlsson Y, Oppegaard O, Skrede S, Itzek A, Wahl AM, Hedetoft M, Bærnthsen NF, Müller R, Nedrebø T (2017b) Association between cytokine response, the LRINEC score and outcome in patients with necrotising soft tissue infection: a multicentre, prospective study. *Sci Rep* 7:42179
- Hansen MB, Rasmussen LS, Garred P, Pilely K, Wahl AM, Perner A, Madsen MB, Hedegaard ER, Simonsen U, Hyldegaard O (2018) Associations of plasma nitrite, L-arginine and asymmetric Dimethylarginine with morbidity and mortality in patients with necrotizing soft tissue infections. *Shock* 49:667–674
- Honore PM, Spapen HD (2016) Pentraxin-3 to better delineate necrotizing soft tissue infection: not really! *Crit Care* 20:173
- Hottelling H (1931) Analysis of a complex of statistical variables into principal components. *J Educ Psychol* 24:417441
- Hsia DYY (1958) Phenylketonuria: the phenylalanine-tyrosine ratio in the detection of the heterozygous carrier. *J Ment Defic Res* 2:8–16
- Ideker T, Krogan NJ (2012) Differential network biology. *Mol Syst Biol* 8:565
- Ioannidis JP, Panagiotou OA (2011) Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses. *JAMA* 305:2200–2210
- Jeong H, Mason S, Barabasi A, Oltvai Z (2001) Lethality and centrality in protein networks. *Nature* 411:41–42
- Jolliffe IT (2002) *Principal component analysis*. Wiley, London
- Jordán F, Nguyen T-P, Liu W-C (2012) Studying protein–protein interaction networks: a systems view on diseases. *Brief Funct Genomics* 11:497–504
- Karakitsou E, Foguet C, de Atauri P, Kultima K, Khoonsari PE, Martins Dos Santos VAP, Saccenti E, Rosato A, Cascante M (2019) Metabolomics in systems medicine: an overview of methods and applications. *Curr Opin Syst Biol* 15:91–99
- Kell DB (2004) Metabolomics and systems biology: making sense of the soup. *Curr Opin Microbiol* 7:296–307
- Kotb M, Norrby-Teglund A, Mcgeer A, El-Sherbini H, Dorak MT, Khurshid A, Green K, Peeples J, Wade J, Thomson G (2002) An immunogenetic and molecular basis for differences in outcomes of invasive group A streptococcal infections. *Nat Med* 8:1398
- Kotb M, Norrby-Teglund A, Mcgeer A, Green K, Low D (2003) Association of human leukocyte antigen with outcomes of infectious diseases: the streptococcal experience. *Scand J Infect Dis* 35:665–669
- Lê Cao K-A, Rossouw D, Robert-Granié C, Besse P (2008) A sparse PLS for variable selection when integrating omics data. *Stat Appl Genet Mol Biol* 7:1
- Le Gall J-R, Lemeshow S, Saulnier F (1993) A new simplified acute physiology score (SAPS II) based on a European/north American multicenter study. *JAMA* 270:2957–2963
- Madsen MB, Skrede S, Bruun T, Arnell P, Rosén A, Nekludov M, Karlsson Y, Bergey F, Saccenti E, Martins Dos Santos VAP, Perner A, Norrby-Teglund A, Hyldegaard O (2018) Necrotizing soft tissue infections—a multicentre, prospective observational study (INFECT): protocol and statistical analysis plan. *Acta Anaesthesiol Scand* 62:272–279
- Madsen MB, Skrede S, Perner A, Arnell P, Nekludov M, Bruun T, Karlsson Y, Hansen MB, Polzik P, Hedetoft M (2019) Patient’s characteristics and outcomes in necrotising soft-tissue infections: results from a Scandinavian, multicentre, prospective cohort study. *Intensive Care Med* 45:1241–1251
- Mehta S, Shelling A, Muthukaruppan A, Lasham A, Blenkinsop C, Laking G, Print C (2010) Predictive and prognostic molecular markers for cancer medicine. *Therapeut Adv Med Oncol* 2:125–148
- Mor G, Visintin I, Lai Y, Zhao H, Schwartz P, Rutherford T, Yue L, Bray-Ward P, Ward DC (2005) Serum protein markers for early detection of ovarian cancer. *Proc Natl Acad Sci U S A* 102:7677–7682
- Norrby-Teglund A, Kotb M (2000) Host-microbe interactions in the pathogenesis of invasive group A streptococcal infections. *J Med Microbiol* 49:849
- Norrby-Teglund A, Chatellier S, Low DE, Mcgeer A, Green K, Kotb M (2000) Host variation in cytokine responses to superantigens determine the severity of invasive group A streptococcal infection. *Eur J Immunol* 30:3247–3255
- Oltvai ZN, Barabási A-L (2002) Life’s complexity pyramid. *Science* 298:763–764
- Pearson K (1901) *On lines and planes of closest fit to systems of points in space*. London, Edinburgh, Dublin *Philos Mag J Sci* 2:559–572
- Petersen A-K, Krumsiek J, Wägele B, Theis FJ, Wichmann H-E, Gieger C, Suhre K (2012) On the hypothesis-free testing of metabolite ratios in genome-wide and metabolome-wide association studies. *BMC Bioinform* 13:120

- Picotti P, Aebersold R (2012) Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat Method* 9:555
- Quezada H, Guzmán-Ortiz AL, Díaz-Sánchez H, Valle-Rios R, Aguirre-Hernández J (2017) Omics-based biomarkers: current status and potential use in the clinic. *Bol Med Hosp Infant Mex* 74:219–226
- Rosato A, Tenori L, Cascante M, de Atauri Carulla PR, Martins Dos Santos VAP, Saccenti E (2018) From correlation to causation: analysis of metabolomics data using systems biology approaches. *Metabolomics* 14:37
- Saccenti E, Hoefsloot HC, Smilde AK, Westerhuis JA, Hendriks MM (2014a) Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics* 10:361–374
- Saccenti E, Suarez-Diez M, Luchinat C, Santucci C, Tenori L (2014b) Probabilistic networks of blood metabolites in healthy subjects as indicators of latent cardiovascular risk. *J Proteome Res* 14:1101–1111
- Saccenti E, Smilde AK, Camacho J (2018) Group-wise ANOVA simultaneous component analysis for designed omics experiments. *Metabolomics* 14:73
- Schwartz S, Kightlinger E, de Virgilio C, de Virgilio M, Kaji A, Neville A, Bennion R (2013) Predictors of mortality and limb loss in necrotizing soft tissue infections. *Am Surg* 79:1102–1105
- Shahzad K, Loor JJ (2012) Application of top-down and bottom-up systems approaches in ruminant physiology and metabolism. *Curr Genomics* 13:379–394
- Simon R (2011) Genomic biomarkers in predictive medicine. An interim analysis. *EMBO Mol Med* 3:429–435
- Simon R (2014) Biomarker based clinical trial design. *Chin Clin Oncol* 3:3
- Staaf J, Glodzik D, Bosch A et al (2019) Whole-genome sequencing of triple-negative breast cancers in a population-based clinical study. *Nat Med* 25:1526–1533. <https://doi.org/10.1038/s41591-019-0582-4>
- Stevens DL, Bryant AE (2017) Necrotizing soft-tissue infections. *N Engl J Med* 377:2253–2265
- Sung J, Wang Y, Chandrasekaran S, Witten DM, Price ND (2012) Molecular signatures from omics data: from chaos to consensus. *Biotechnol J* 7:946–957
- Thänert R, Itzek A, Hoßmann J, Hamisch D, Madsen MB, Hyldegaard O, Skrede S, Bruun T, Norrby-Teglund A, Medina E (2019) Molecular profiling of tissue biopsies reveals unique signatures associated with streptococcal necrotizing soft tissue infections. *Nat Commun* 10:1–15
- Vafae F, Diakos C, Kirschner MB, Reid G, Michael MZ, Horvath LG, Alinejad-Rokny H, Cheng ZJ, Kuncic Z, Clarke S (2018) A data-driven, knowledge-based approach to biomarker discovery: application to circulating microRNA markers of colorectal cancer prognosis. *NPJ Syst Biol Appl* 4:1–12
- Van 't Veer LJ, Dai H, Van de Vijver MJ, He YD, Hart AAM, Mao M, Peterse HL, Van Der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415:530–536
- Van de Vijver MJ, He YD, Van 't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ (2002) A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347:1999–2009
- Vignoli A, Tenori L, Luchinat C, Saccenti E (2018) Age and sex effects on plasma metabolite association networks in healthy subjects. *J Proteome Res* 17:97–107
- Vignoli A, Tenori L, Giusti B, Valente S, Carrabba N, Balzi D, Barchielli A, Marchionni N, Gensini GF, Marcucci R, Gori AM, Luchinat C, Saccenti E (2020) Differential network analysis reveals metabolic determinants associated with mortality in acute myocardial infarction patients and suggests potential mechanisms underlying different clinical scores used to predict death. *J Proteome Res* 19:949–961
- Villoslada P, Baranzini S (2012) Data integration and systems biology approaches for biomarker discovery: challenges and opportunities for multiple sclerosis. *J Neuroimmunol* 248:58–65
- Vincent JL, Moreno R, Takala J, Willatts S, de Mendonça A, Bruining H, Reinhart CK, Suter PM, Thijs LG (1996) The SOFA (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. On behalf of the working group on sepsis-related problems of the European Society of Intensive Care Medicine. *Intensive Care Med* 22:707–710
- Wiśniewski JR, Mann M (2012) Consecutive proteolytic digestion in an enzyme reactor increases depth of proteomic and phosphoproteomic analysis. *Anal Chem* 84:2631–2637
- Wold S, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. *Chemom Intell Lab Syst* 58:109–130
- Wong C-H, Khin L-W, Heng K-S, Tan K-C, Low C-O (2004) The LRINEC (laboratory risk indicator for necrotizing fasciitis) score: a tool for distinguishing necrotizing fasciitis from other soft tissue infections. *Crit Care Med* 32:1535–1541