Full Review

Systems biology: Building a useful model from multiple markers and profiles

Paul Mayer¹, Bernd Mayer¹,² and Gert Mayer³

¹Emergentec Biodevelopment GmbH, Vienna, Austria, ²Institute for Theoretical Chemistry, University of Vienna, Vienna, Austria and ³Department of Internal Medicine IV (Nephrology and Hypertension), Medical University Innsbruck, Innsbruck, Tyrol, Austria

Correspondence and offprint requests to: Gert Mayer; E-mail: gert.mayer@i-med.ac.at.

Abstract

The pathophysiology of diabetic nephropathy (DN) is driven by a complex, multi-faceted interplay of numerous molecular processes (protective as well as damaging) and the balance between these, rather than the activity of a single pathway, determines clinical presentation and outcome. We present a concept for deriving a biomarker panel aimed to represent the relevant processes involved. Our approach rests on a hybrid gene/protein interaction network that holds ample information on molecular features (nodes) and their relations (edges), as a result providing a basic structure to navigate in molecular content and context being identified as relevant in DN. Extensive literature search on omics studies in DN provided a molecular feature list mapping to a total of 2175 unique protein-coding genes (13 from single nucleotide polymorphisms (SNPs), 12 as targets from relevant miRNAs, 1583 from transcriptomics, 5 from proteomics and 53 from metabolomics via linking to enzymes; 509 features were identified from multiple sources). Two hundred and eighty-seven further human protein-coding genes associated with DN were derived from searching NCBI Pubmed (utilizing MeSH and gene-to-pubmed). Text mining of patents and clinical trial descriptions in the context of DN further added about 1 000 features. These data were used to label the respective nodes in the interaction network, as a result obtaining a DN-specific subgraph. Application of a segmentation algorithm on this subgraph allowed the identification of DN-specific molecular units, each characterizing a cluster of genes/proteins with a high internal functional association. We interpret each such unit as a functionally relevant molecular process contributing to the presentation of DN, and the total set of such units as a molecular model of DN. We propose that selecting appropriate biomarkers from each unit might allow the description of a patient’s specific ‘type’ of DN, ultimately leading to a better stratification of patients regarding progression risk and optimal interventional approach.

Keywords: biomarker; diabetic nephropathy; data integration; disease model; omics

In 2003, 7.8% of the adult European population suffered from diabetes mellitus, a number projected to rise to 9.1% (58.6 million) by 2025 [1]. In the USA, the prevalence of diabetic nephropathy (DN), defined as albuminuria >30 mg/g creatinine and/or an estimated glomerular filtration rate (eGFR) <60 mL/min/1.73 m², increased from 2.2% in 1988 to 3.3% in 2008 [2]. Large-scale epidemiological European data on non-dialysis-dependent diabetic chronic kidney disease (CKD) stages I–V are not available, but the incidence of end-stage renal disease doubled from 13.6 pmp in 1992 to 27.6 pmp in 2005. Despite a 5-year mortality rate of 67%, the prevalence increased from 46.4 to 113.1 pmp within the same time period [3]. Even though optimized metabolic control and/or reduction of blood pressure, preferably by blocking the renin–angiotensin–aldosterone system, improves the prognosis (the efficacy being dependent on the CKD stage and definition of progression by change in albuminuria and/or eGFR) in type 1 and type 2 disease [4–8], these interventions are not without side effects [9] and progression still occurs in large groups of patients once the disease is clinically evident. These individuals might benefit from intensified standard medical therapy started at even earlier stages, when the intrinsic renoprotective/regenerative potential is better preserved, or may be approached by utilizing drugs addressing novel targets, or by repositioning of medications currently used to treat other diseases. In this work, we discuss strategies brought forward in the realm of systems biology aimed at building molecular models of DN, and from there delineating biomarker panels allowing a patient-specific assessment of DN status for supporting treatment decisions.

In order to reach such a goal, we need an improved understanding of the pathophysiology of DN in the first place. It is evident that in type 2 diabetes, CKD progression can occur in the presence or absence of albuminuria [10], but DN is probably a much more complex and diverse phenotype with a multi-faceted interplay of numerous molecular processes (protective as well as damaging, each affected by individual genetic susceptibility, environment
and treatment), and the balance between these, rather than the activity/disturbance of a single pathway, determines clinical presentation, prognosis, treatment response and consequently outcome. Deciphering the balance of renoprotective and harmful processes at an early stage of DN appears particularly interesting when aiming at novel treatments and ultimately preventive approaches, as such not being addressed therapeutically so far.

When confronted with diversity, biomedical research has traditionally pursued a reductionist approach, dividing a problem into smaller pieces (e.g. ‘what is the role of the RAAS in DN?’) under the assumption that studying less complex, predefined subunits of the disease will provide more definitive answers. Consequently, however, the ability to assess the quality and quantity of interactions between different pathophysiological processes or treatments is limited. Actually a randomized, controlled trial, the gold standard in interventional clinical research, tests a single hypothesis and even a multifactorial design at best allows post-hoc analyses for interactions, but with inherent limitations [11].

Consequently in order to advance, DN needs to be addressed on a more basic level [12], and here we propose the following workflow:

- determine a comprehensive set of molecular features which are associated with DN,
- on the basis of this set of individual molecular features, determine the associated set of molecular processes reflecting the pathophysiology of DN,
- for each identified process, derive a representative biomarker (resulting in a panel of markers) allowing quantitative assessment of each individual process,
- use this biomarker panel to characterize the individual patient,
- based on the readout, determine the patient-specific ‘type’ of DN and associated progression risk,
- use this knowledge to address the specific types of DN with specific therapeutic interventions.

In order to deliver on these promises, large-scale molecular characterization of clinically relevant presentations [e.g. micro- or macroalbuminuria, preserved or reduced glomerular filtration rate (GFR)] and outcomes (remission, stable disease, slow or fast progression, beneficial treatment response or resistance to specific therapy, etc.) of DN is needed. During the last decades, high throughput explorative, hypothesis-generating methods potentially serving these claims have been developed fostered by emerging multiplexed assay technology, ultimately leading to the omics revolution. It was only in 1993, that Fodor et al. published the first data resting on oligonucleotide chips [13], and similar platforms have been developed to study, within a single experiment, thousands of transcripts (transcriptomics), proteins (proteomics), metabolites (metabolomics) or single nucleotide polymorphisms (SNPs). Whereas genomic technologies localize SNPs throughout the genome, transcriptomics uses microarrays or next generation sequencing, originally for the (semi-)quantitative assessment of the expression of ∼20 000 human protein-coding genes, now further expanding into the large space of regulatory, non-coding (nc)RNAs. In case compartment/cell specific expression profiling is desired in complex tissues, laser capture microdissection can be applied up front. Modern proteomic profiling utilizes fractionation techniques such as capillary electrophoresis coupled to mass spectrometry to study the abundance of proteins and protein fragments in blood, urine or tissues. A similar technology is pursued in metabolomics to study the concentration of low-molecular weight compounds, being a representation of enzymatic activities as well as environmental impact.

Certainly, omics still faces limitations including incompleteness of molecular catalogues (with ncRNAs as an example being still not fully covered), further expansion on standardization in the context of clinical samples (detailed characterization of phenotypes, omics technology-specific sample collection, standards in experimental processing) [14] and fundamentally the fact that omics allows in the first place only a description of a specific molecular state for a specific sample—per se not providing insights into any information on molecular processes and their interaction dynamics [15]. Despite these boundaries, significant omics profiling has been brought forward specifically on the later stages of DN (in the recent past seeing initiatives for complementing also on early stages), providing the ground for integrative analysis approaches focusing on deciphering a more complete picture of the molecular processes of the different stages of DN. Knowledge of such a process landscape allows the assessment of a patient’s individual molecular signature of the specific disease state, from there supporting optimized treatment. Concepts and methodologies for realizing such strategies have been developed in the field of Systems Biology [16], now also traversing into translational clinical research (Systems/Pathway/Network Medicine) [17], aiming at patient stratification (and ultimately personalization) on the basis of a more holistic level by integrating clinical descriptors with extended molecular profiling [18].

A recent review by He et al. summarizes efforts in systems biology in the realm of kidney diseases [19], discussing omics data integration in the context of improving early detection, predicting disease progression and monitoring treatment response. The kidney is organized as a highly orchestrated interplay between various levels of molecular organization reaching from the genome to the metabolome, which also requires respect for the various cell types and compartments. On a specific molecular level, integrative analysis with a phenotype of interest is already standard, as exemplified for genome-wide association studies, where the combination of SNPs at multiple loci provided an association with the GFR in DN, although individual SNPs showed only weak effects [20]. However, the interactive nature of the various levels of molecular functions in the context of the phenotype cannot be covered by this approach. Here, a fundamental expansion next to statistical analysis (e.g. significance of associations) comes into play: transforming findings from omics profiling, literature search as well as from hypothesis-driven approaches to a biological (network) context.

Molecular interaction networks aim at providing a representation of molecular features (e.g. proteins, protein-coding genes or metabolites, presented as nodes) together with their interactions (e.g. protein–protein interaction,
protein–substrate binding, presented as edges). Types of such molecular interaction networks are manifold, differing in scope (nature of interactions represented, e.g. focusing explicitly on transcriptional control via transcription factors, or specifically on enzyme–substrate interactions for representing metabolic processes), type of molecular entities included (genes, proteins and metabolites), coverage (e.g. how many of the protein-coding genes are effectively represented) and the level of evidence necessary for inclusion (i.e. how well established is the experimental background on an interaction) [21]. The Kyoto Encyclopedia of Genes and Genomes (KEGG) can serve as an example, displaying at present ∼150 well-defined human molecular core pathways (e.g. apoptosis cascade, renin–angiotensin pathway), currently holding approximately one-third of all protein-coding genes, further complemented by disease-associated pathways (e.g. pathways of cancer). Molecular features identified as relevant, for example in a specific transcriptomics study (or in any other specific omics profiling) on DN, can now be mapped onto the KEGG in order to decipher in which pathways such features are of relevance. Consequently, information on relevant pathways is retrieved by utilizing lists holding relevant molecular features. Although the KEGG and other similar databases are valuable tools, the very specific nature of such repositories introduces a bias (KEGG looking only at a specific set of pathways) and additionally these ontologies are mostly not disease specific: regardless of the clinical phenotype, the experimental results are always projected on the same pathways. Recently attempts have been started for generating broad scope (‘all types of interactions) and broad coverage (‘all’ molecular feature levels) interaction networks [22], for selected examples already being used for studying kidney diseases [23]. Such networks aim at encapsulating the various types of molecular interactions including protein–protein, protein–substrate, protein–DNA or even further genetic interactions, with synthetic lethality as the most prominent example [24]. Clearly, completeness of molecular catalogues is an issue here, where omics technologies have brought us significant expansions of knowledge on molecular entities. Next to appropriate molecular entity referencing via provision of specific identifier sets regarding human genes, transcripts, proteins and metabolites as handled by major repositories such as ENSEMBL or NCBI, also further functional and context annotation (molecular interactions, assignment to specific functional pathways or data on tissue-specific expression as e.g. handled in the Genecards database [25]) of molecular entities has significantly expanded. Next to more generic information, also organ-specific or disease-specific information has accumulated, with the renal gene ontology initiative driven by the European Bioinformatics Institute (EBI) [26] as an example. Further sources available for annotation of molecular features include e.g. drugs linked to their targets (as provided in Drugbank [27]), or linking molecular features with diseases utilizing Pubmed MeSH (Medical Subject Headings). Following these lines, a rich data space on molecular features, each characterized on multiple functional levels, has emerged.

Next to a best possible characterization of molecular features, maximum coverage of their interactions (or more general relations) is essential, culminating in the question, ‘how many such relations are there’ [28]? Consolidating interaction databases for human protein-coding genes/gene products resulted in ∼100 000 interactions. Hybrid networks, in part further expanded by inferred (predicted) interactions significantly exceed this number, certainly afflicted with an increasing number of false-positive interactions. As an example, inclusion of more complex types such as genetic interactions needs traversing interaction information on cross-species, with the majority of data available only for yeast [29].

As with the molecular feature space, also the various datasets on interactions have seen significant expansions and updates, altogether being a result of ‘interactomics’ screening [30].

The consolidation of information on molecular features and their specific interactions in hybrid network structures (Figure 1) results in composite relation networks covering the various levels of molecular organization, as a result providing the basic structure for navigation in molecular content and context. Examples of such hybrid networks are omicsNET [22] or STRING [31].

The next step is to add information on disease-specific molecular changes (in our case regarding DN), and from there delineating a molecular disease model.

Multilevel omics data of DN (Table 1) provide us with individual molecular feature lists mapping to a total of 2175 unique protein-coding genes (13 from SNPs, 12 as targets from miRNAs, 1583 from transcriptomics, 5 from proteomics and 53 from metabolomics via linking to enzymes; 509 features are identified from multiple sources).

These omics datasets resulted in the evidence for e.g. pathophysiologically significant sequence variants of uromodulin [32], variants in the promoter modules of transcription factor NF-kB-dependent genes in the context of stress response [33], specific miRNAs as miRNA-21 potentially playing a role in tubular protection [34], differential expression levels of vascular endothelial growth factor -A comparing stable and progressive renal disease as identified on the level of mRNA [35], proteomics profiles holding collagen fragments specific for DN progression [36] and esterified/non-esterified fatty acids specific in DN [37].

Next to consolidating omics studies, 287 human protein-coding genes can be derived from searching NCBI Pubmed (utilizing MeSH and gene-to-pubmed). Text mining of patents and clinical trial descriptors in the context of DN further adds about 1000 additional features. Next to redundancy in these datasets (increasing the evidence for robust association for a fraction of features), many of these associations may only reflect specific peculiarities of a specific study design, a large fraction may not be disease specific (or even artifacts), and only a subset will actually reflect the cascade of molecular processes intrinsically characterizing DN [38].

Nonetheless, the obvious next question is how to integrate the heterogeneous feature sets into an interaction network for ultimately describing disease-specific molecular processes. Technically, this is achieved by bringing the features to the name space also embedded in the interaction network to be used (e.g. human protein-coding genes/proteins), followed by the labelling of all nodes of the network which are relevant in the disease (Figure 2).
With this approach, relevance is defined on the level of the individual omics studies, i.e. utilizing a specific study design in the scope of the clinical question, applying a specific experimental technology for deriving omics profiles, followed by utilizing specific statistical test procedures for extracting the molecular feature set showing a statistically significant association with the phenotype.

Only such relevant features from such a statistical perspective are then further considered on the network level.

The second step is then the identification of disease-specific molecular processes on the basis of the disease-specific network annotation, i.e. what are the functional molecular units (groups of features with distinct relations) being specifically affected by the disease. A frequently applied procedure is to use interaction networks already being categorized into functional units. KEGG pathways are an example, and using gene set enrichment analysis [38] for identifying those processes being significantly affected from a statistical perspective (i.e. computing an enrichment value taking into account the total number of molecular players relevant in a specific process and how many of these are also identified as affected in the disease according to the omics profiles). However, as mentioned above, such approaches show shortcomings as limited coverage of the human protein-coding gene universe, but even more importantly, such networks are representing specific types of interactions (as KEGG mainly representing interactions within a specific process) and per se do not represent disease-specific details. This is why hybrid networks may improve the identification of disease-specific processes, on the one hand, by increasing coverage of molecular entity representation, on the other hand, by encoding a more general view of the numerous types of relations, which may resemble the types of relations which are important in the disease-specific pathophysiology. The task now is to identify specific functional

Table 1. Reference, omics type, sample matrix and CKD stage for selected omics studies of relevance in the context of DN

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Omics type</th>
<th>Tissue</th>
<th>DN stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baelde et al. 2004 [49]</td>
<td>Transcriptomics</td>
<td>Glomerular</td>
<td>Early stage</td>
</tr>
<tr>
<td>Berthier et al., 2009 [50]</td>
<td>Transcriptomics</td>
<td>Glomerular</td>
<td>Early stage</td>
</tr>
<tr>
<td>Woroniecka et al., 2011 [51]</td>
<td>Transcriptomics</td>
<td>Glomerular</td>
<td>Later stage</td>
</tr>
<tr>
<td>Berthier et al., 2009 [50]</td>
<td>Transcriptomics</td>
<td>Tubulointerstitium</td>
<td>Later stage</td>
</tr>
<tr>
<td>Cohen et al. 2008 [52]</td>
<td>Transcriptomics</td>
<td>Tubulointerstitium</td>
<td>Later stage</td>
</tr>
<tr>
<td>Schmid et al., 2006 [53]</td>
<td>Transcriptomics</td>
<td>Tubulointerstitium</td>
<td>Later stage</td>
</tr>
<tr>
<td>Woroniecka et al. 2011 [51]</td>
<td>Transcriptomics</td>
<td>Tubulointerstitium</td>
<td>Later stage</td>
</tr>
<tr>
<td>Chambers et al. 2010 [54]</td>
<td>SNPs</td>
<td>Whole blood</td>
<td>–</td>
</tr>
<tr>
<td>Köttingen et al. 2010 [33]</td>
<td>SNPs</td>
<td>Whole blood</td>
<td>–</td>
</tr>
<tr>
<td>Lim et al. 2009 [55]</td>
<td>SNPs</td>
<td>Whole blood</td>
<td>–</td>
</tr>
<tr>
<td>Krupa et al. 2010 [56]</td>
<td>miRNAs</td>
<td>Whole kidney</td>
<td>Later stage</td>
</tr>
<tr>
<td>Han et al. 2011 [42]</td>
<td>Metabolomics</td>
<td>Plasma</td>
<td>Later stage</td>
</tr>
<tr>
<td>Kim et al. 2007 [57]</td>
<td>Proteomics</td>
<td>Serum</td>
<td>Various</td>
</tr>
<tr>
<td>Jain et al. 2005 [58]</td>
<td>Proteomics</td>
<td>Urine</td>
<td>Early stage</td>
</tr>
<tr>
<td>Mischak et al. 2004 [59]</td>
<td>Proteomics</td>
<td>Urine</td>
<td>Various</td>
</tr>
<tr>
<td>Dihazi et al. 2007 [60]</td>
<td>Proteomics</td>
<td>Urine</td>
<td>Various</td>
</tr>
<tr>
<td>Sharma et al. 2005 [61]</td>
<td>Proteomics</td>
<td>Urine</td>
<td>Later stage</td>
</tr>
<tr>
<td>Rossing et al. 2008 [62]</td>
<td>Proteomics</td>
<td>Urine</td>
<td>Later stage</td>
</tr>
<tr>
<td>Alkhalaf et al. 2010 [63]</td>
<td>Proteomics</td>
<td>Urine</td>
<td>Later stage</td>
</tr>
</tbody>
</table>

Downloaded from https://academic.oup.com/ndt/article-abstract/27/11/3995/1815171/Systems-biology-building-a-useful-model-from by guest on 15 September 2017
segments on a hybrid interaction network populated by disease-specific molecular signatures. Next to fundamental questions on what the specific properties of disease-specific functional units are (as recently reviewed by Barabasi et al. [39]) algorithmic challenges on effective segmentation are manifold, with the ‘mcode’ algorithm [40] being one approach. The background of this segmentation approach is to detect densely connected regions in large-scale interaction networks, and to interpret such regions as functionally distinct entities. Applying such algorithms on disease-specific interaction networks (i.e. only considering the part of the entire network holding nodes being associated with the disease) consequently provide us with disease-specific functional units.

At this point, multilevel omics data on the specific clinical phenotype (as listed in Table 1) are integrated with the aim of representing the specific molecular feature state space encapsulating genetic predisposition (via SNPs) and the environmental impact assessed on transcriptional control, protein effector level and metabolomic readout. This consolidated molecular feature space is mapped on a hybrid molecular relations network providing improved coverage of the molecular catalogue, together with broad representation of relations between the molecular entities. On the basis of the disease-specific network, graph segmentation is performed to identify functional units, where each such unit itself represents a set of molecular features. On this basis, the level of the relation between the units can be calculated, representing the aggregate relation as sum of relations between individual nodes (features) composing the individual units. Delineating units rests on specific properties of nodes embedded in such units, whereas computing the relation between units allows hypotheses to be built regarding the interplay of the units as such. Importantly, such units are disease specific and not constant reference elements (as e.g. KEGG pathways), promising identification of novelty, however, at the cost of evidence regarding true functional context. Populating the hybrid network separately with features identified at the early stage and at the later stage of DN results in different disease-specific subgraphs (depending on the molecular diversity accompanying the disease state), from there leading to different units of which some may become apparent at the early stage (e.g. predominantly renoprotein-associated), being replaced by damage-associated units at the later stages.

We consider such representation, as shown in Figure 3B for DN, as a model of disease composed of multiple markers and profiles (holding 691 of the total 2175 features derived from the omics studies), providing a comprehensive picture of the molecular processes involved, with the number of features embedded in the units ranging from 10 to 194, with 51 on average. Having such a model in hand allows selection of the biomarker candidates specifically probing the set of molecular units (i.e. each being centrally embedded in a specific functional unit). As a result, a biomarker panel is generated aiming at probing the entire molecular landscape relevant in DN (in contrast to just a single such process when using a single marker), but also allowing to test the temporal importance of specific units and unit combinations in the course of disease progression. Mining Pubmed MeSH for genes reported as biomarker candidates in the context of DN provides 93 such gene/protein features (disease term ‘diabetic nephropathy’ and ‘biomarker’, PubMed data status as of April 2012). A total of 14 of these are also members of the DN units model shown in Figure 3B, consequently providing a valuable start set. Equally important, the selection of new biomarker candidates covering units presently not addressed on this level becomes straightforward (and by this reaching a more complete molecular characterization of the disease on the level of a multimarker panel).

A recent review on reported biomarker candidates for DN progression clearly identified shortcomings of individual markers regarding the assessment of the variance of progression on the individual patient level [41]. Presumably, the specific signature of combining such markers will provide better conclusions on the individual patient level, as reviewed for proteomics techniques in [42]. Quantification of such biomarker panels on the individual patient level may allow us to assign the relevant unit combinations to specific patients, in turn allowing the identification of the impact of medication on the unit assignment (reflected by changes in the biomarker profile).
but also allowing the study of the evolution of the patient-specific molecular disease state over time in a longitudinal assessment of the multimarker panel. Multiplexed assay technologies are at hand for bringing such an approach to the clinical setting, promising patient stratification based on unit composition at early stages (reflecting the individual molecular background at disease onset measured by a multimarker panel), and from there delineating the likely molecular trajectory the disease will take, potentially reflecting the change from dominating renoprotective to damage-associated processes. This information in turn is the prerequisite for tailored therapy.

The approach presented in this work focuses on deriving a biomarker panel regarding a specific clinical phenotype and utilizing omics profiles from cross-sectional analyses. However, such approaches may be taken even further focusing on the development of disease states in a time-dependent fashion already at the omics profiling level, as recently presented by Chen et al. [43]. This study utilized multi-level omics profiling on an individual subject level in a longitudinal fashion for delineating integrative personal omics profiles associated with the development of disease (in contrast to biomarker panels as presented in our approach).

As noted above, multisource integration further allows drug targets and drugs to be linked to the molecular features composing the relations network, consequently providing us with unit-specific candidate medication. Such an approach may support decision making on which patients appear to be better responders to e.g. RAAS blockade, or if there are other medications also addressing DN-specific units (repositioning), or if there are novel targets embedded into such units [12].

**Outlook**

Conceptual strategies and methodological implementations for the identification of a DN specific-molecular process landscape have become available, followed by utilizing such approaches in multimarker panel definition, which after validation in patient cohorts promises support in patient stratification enabling tailored treatment (Figure 4).

However, in order to truly bring such efforts forward, disease-specific platforms need to be put into place holding cross-omics profiles together with extended sample and clinical phenotype descriptions. Such a basis will allow us to derive stage, compartment and eventually also cell type-specific units, thereby expanding the model. Valuable data collections for kidney research covering transcriptomics and proteomics include the nephromine platform [44] and the KUPKB database [45], also providing extensive search, data comparison, functional annotation as well as network visualization tools rendering the platforms as effective research tools. Such data collections serve as a source for efforts in the realm of systems biology analyses for...
delineating disease models, as proposed in this work, and data management and integration concepts for feeding network analysis, to be put into place [46].

Of further importance is the validation of any such hypothesis derived from disease models, both on a functional level utilizing proper in vitro and in vivo models of DN, and most essential also in the clinical context, here requiring massive biobanking efforts, for ultimately evaluating the clinical impact of such findings in the realm of epidemiological boundaries of the disease [47].

Next to these promises, the following limits and challenges need to be kept in mind at the present stage:

- we are still confronted with limited knowledge on molecular catalogues, but particularly on molecular interactions,
- at present molecular interaction networks only provide a first approximation of the true interactome,
- optimal concepts and algorithms for extracting disease-specific units are still lacking,
- for most of the clinical phenotypes, omics profiling on the multiple levels is far from being complete.

Although significant data are available in the public domain, gaps are evident, specifically for the earlier stages of DN. Furthermore, such models only serve for hypothesis generation, consequently demanding validation on the pathophysiological and clinical level. For such integrative approaches as outlined in Figure 4, large-scale research consortia such as SysKid [48] provide platforms with all the necessary tools at hand for coping with the promises of bringing a more individualized therapy of DN forward.

Acknowledgement. We are thankful to Marcus Saemann MD (Medical University Vienna) and Peter Schratzberger (Medical University Innsbruck) for critically reviewing the manuscript and to Raul Fechete, Andreas Heinzl and Paul Perco for providing their expertise in bioinformatics and Systems Biology.

Funding. The research leading to these results has received funding from the European Community’s Seventh Framework Programme under the grant agreement No. 241544.

Conflict of interest statement. The authors confirm that there are no conflicts of interest. The authors declare that the results presented in this paper have not been published previously in whole or part, except in abstract format.

References
3. http://www.eugloreh.it
27. http://www.drugbank.ca/
44. http://www.nephromine.org
45. http://www.kuplb.org
46. Lukas A, Mayer B. Data annotation and relations modeling for integrated Omics in clinical research. JOAB J 2010; 1: 15–23

Received for publication: 26.5.2012; Accepted in revised form: 10.9.2012