



scientific report

Structure homology and interaction redundancy for discovering virus-host protein interactions

Benoît de Chassey ¹, Laurène Meyniel-Schicklin ^{2,3}, Anne Aublin-Gex ^{2,3}, Vincent Navratil ^{2,3,†}, Thibaut Chantier ^{2,3}, Patrice André ^{1,2,3} & Vincent Lotteau ^{2,3+}

¹Hospices Civils de Lyon, Hôpital de la Croix-Rousse, Laboratoire de virologie, Lyon, ²Inserm, U1111, Lyon, and ³Université de Lyon, Lyon, France

Virus-host interactomes are instrumental to understand global perturbations of cellular functions induced by infection and discover new therapies. The construction of such interactomes is, however, technically challenging and time consuming. Here we describe an original method for the prediction of high-confidence interactions between viral and human proteins through a combination of structure and high-quality interactome data. Validation was performed for the NS1 protein of the influenza virus, which led to the identification of new host factors that control viral replication.

Keywords: interactome; prediction; protein interaction; structure; virus

EMBO reports advance online publication 6 September 2013; doi:10.1038/embor.2013.130

INTRODUCTION

Viruses are obligate parasites that rely on cellular functions to replicate. Identifying the cellular functions that a virus must use, counteract or interfere with to assure its replication is therefore a major issue for both basic knowledge and drug discovery. Biological processes being largely dependent on protein–protein interactions, cellular functions involved in viral replication can be identified by mapping virus–host protein–protein interactions and by constructing virus–host interactomes. Until 2007, viral protein–host protein interactions have essentially been identified in low-throughput experiments. Several databases are integrating these data spread in the literature [1–2]. The construction of physical viral ORFeomes and the development of high-throughput technologies, such as yeast two-hybrid (Y2H) and tandem affinity

Received 14 February 2013; revised 29 July 2013; accepted 31 July 2013; published online 6 September 2013

purification coupled to mass spectrometry, led to the publication of the first virus-host interactomes [3-4]. Although incorporation of data from different methods or variation of the same method [5] has improved the quality of the data sets, diversification of methods is still clearly needed to generate high-quality comprehensive virus-host interactomes. In addition, regarding the size of host genome and the huge diversity of viruses, millions of binary interactions remain to be tested. Therefore, accurate and rapid methods for the identification of cellular interactors controlling viral replication is a major issue and a crucial step towards the selection of original therapeutic targets and drug development. This could benefit from predictive methods preselecting subsets of putative interactions. Computational methods are essentially dedicated to predict intraspecies interactions [6] but applications to pathogen-host protein interactions are now emerging [7]. These methods are generally evaluated by the overlap between predicted and published interaction data sets. Here, we present and experimentally validate an original method for the prediction of virus-host interactions combining protein structure homology and interaction redundancy.

RESULTS AND DISCUSSION Principle of the method

Several methods have proposed to use structural information to predict protein–protein interactions [8–10]. The method described here relies on the assumption that when two proteins are structurally homologous, they are more likely to have interactors in common (Fig 1A). Using the Structural Classification of Proteins from SCOP database [11] and the high-quality data sets of protein–protein interactions from the VirHostNet database [2], we showed that this assumption is a general feature of human and viral proteins (supplementary information online). Briefly, for a viral protein having a solved structure, structurally homologous human and viral proteins are first selected. Interactors of these homologous proteins are then identified. These proteins are considered as putative interactors and ranked according to a score that favors proteins independently identified from multiple structural homologues.

¹Hospices Civils de Lyon, Hôpital de la Croix-Rousse, Laboratoire de virologie, 103 grande rue de la Croix-Rousse, Lyon, F-69004,

²Inserm, U1111, 21 Avenue Tony Garnier, Lyon, F-69007,

³Université de Lyon, Lyon, France

[†]Present address: Laboratoire de Biométrie et Biologie Évolutive, CNRS, Université Claude Bernard Lyon 1, 43 boulevard du 11 Novembre 1918, 69622 Villeurbanne Cedex, France.

⁺Corresponding author. Tel: +33 437262971; Fax: +33 437282341; E-mail: vincent.lotteau@inserm.fr



Fig 1 | Description of the method. (A) Principle of the method. (B) Detailed steps of the method applied to influenza A virus NS1 protein. The score favors independent predictions provided by interaction redundancies. Score (putative interactor) = $[(n_{\rm H}^2/k_{\rm H}) + (n_{\rm V}^2/k_{\rm V})] \times a$. With $n_{\rm H}$ and $n_{\rm V}$ number of distinct human interactors and viral interactors, respectively; $k_{\rm H}$ and $k_{\rm V}$ degree of the protein in the human protein–protein interaction data set, are a for a prediction from both human and virus–human data sets, a = 2 from only virus–human data set, 1 else. H, human; V, virus.

Prediction of influenza A virus NS1 protein interactions As a proof of concept, we applied this method to identify new partners of the influenza A NS1 protein whose full-length structure has been solved by x-ray crystallography [12], (Fig 1B). A total of 22,052 human protein structures and 4,933 viral protein structures respectively corresponding to 3,163 human proteins and 440 viral proteins were extracted from the Protein Data Bank (PDB). The standalone version 3.1 of the Dali programme was used to perform one-against-all pairwise alignments between NS1 and these large data sets of viral and human protein structures [13]. NS1 structural homologues with a Dali z-score between 2 and 20 were retained to include weakly homologous proteins and exclude fully homologous NS1 proteins [14]. All homologues that are fulllength or part of NS1 proteins of influenza A viruses were also excluded. Hence, known NS1-human interactors were not favored. Influenza NS1 protein was found structurally homologous to 344 human proteins and 42 viral proteins. Using high-quality data sets of protein-protein interactions from the VirHostNet database [2], a total of 1,384 human proteins were predicted to interact with NS1- 1,377 interacting with human proteins that are homologous to NS1 and 86 interacting with viral proteins that are homologous to NS1. These 1,384 proteins were then ranked according to the number and origin of the NS1 homologues weighted by their degree. The 5% top-ranked candidates, corresponding to 69 proteins, were selected for further validation (Fig 1B, supplementary Table S1 online). In silico mapping of the predicted interactions can also be performed when the structure of domains or fragments are available. For NS1, an RNA-binding domain and an effector domain have been structurally defined (PDB identifiers: 2Z0A and 3D6R, supplementary Table S2 online). Fifty eight of the 69 predicted interactors for the full-length NS1 protein were predicted to interact with the RNA-binding domain while none of them were predicted to interact with the effector domain (supplementary Table S1 online).



Fig 2 | Functional characterization of the 26 best-predicted interactors. (A) The physical and genetic interactions have been tested (white) or not (grey). Essential host factors are in green and restriction host factors are in red. (B) BID–NS1 interaction mapping. Interaction of myc–BID with GST–NS1 and GST–NS1 RBD but not with GST alone or GST–NS1 effector domain. (C,D) Impact of silencing indicated genes on viral replication. Values are the number of PFUs normalized to control siRNA. The mean values \pm - s.d. from three independent experiments are shown. (C) Restriction host factors. (D) Essential host factors. GST, glutathione S-transferase; PFU, plaque-forming unit; RBD, RNA-binding domain; siRNA, short interfering RNA.

Functional analysis of NS1 interactors

Sixty-four of the 69 predicted interactors are known to be expressed in the normal lung or trachea in accordance with the tropism of the virus (supplementary Table S1 online). Analysis of enriched gene ontology terms and KEGG pathways corresponding to the 69 predicted interactors of NS1 confirms and provides more molecular supports to the interplay of NS1 with essential cellular functions. A significant proportion of the predicted interactors being involved in apoptosis (Benjamini *P*-value = 6.31×10^{-6}), this new data set might help clarifying the complex and controversial role of NS1 and apoptosis in viral replication [15-18]. In the enriched RIG-I signalling pathway (Benjamini *P*-value = 5.4×10^{-3}), NS1 is known [19,20] to interfere with the signalling of viral RNA recognition by the pathogen recognition receptors by interacting with RIG-I and TRIM25 [21,22]. Here, the interaction of NS1 with 4 IKK proteins further supports its ability to interfere with this signalling at a more downstream step. Predicted interaction of NS1 with nine proteins of the mitogen-activated protein kinase pathway (Benjamini P-value = 1.9×10^{-3}) is also demonstrative of the crucial role of this pathway whose late activation is required for efficient replication of the virus [23].

Validation of NS1 interactors

Experimental validation of the predicted interactions was performed by Y2H and GST pull down for 32 candidates. Eighty-one percent of the interactions were confirmed by Y2H and 72% by co-affinity precipitation (supplementary Fig S1 online). Interestingly, 100% of the interactions were validated by at least one method when the predicted interactor had a

score > = 3. The validation rate was 83% for predicted interactor having a score <3 (supplementary Table S1 online). Therefore, using a score > = 3 as a threshold to select the best interactors, the prediction method allowed the identification of very highconfidence interactors. Accordingly, from the 69 interactors initially predicted, a very high-confidence set of 26 NS1 interactors was retained (Figs 1B and 2A). Performance assessment of the prediction method shows that this score threshold leads to an optimal success regarding specificity and precision (supplementary information online). Twenty-five out of these 26 interactors were new, illustrating the potential of this method to predict novel interactions (supplementary Table S1 online). From the 26 interactors, 24 were predicted to interact with the RNAbinding domain but not the effector domain of NS1 (supplementary Table S1 online). This was confirmed by GST pull down for the NS1-BID interaction further validating the method (Fig 2B).

As the overlap of predicted interactors for full-length NS1 and the RBD is not 100%, both domains and full-length proteins are essential to explore the full potential of this method.

Impact of silencing NS1 interactors on viral replication

The 26 high-confidence NS1 predicted interactors were further tested for their ability to control influenza virus replication. Systematic gene silencing was performed with three independent small interfering RNA per gene and virus production was first evaluated 24 h post infection by measuring neuraminidase activity in the culture supernatant. When viral production was inhibited or enhanced by at least twofold with at least two short interfering RNAs (siRNAs) in at least two out of three experiments, viral production was further quantified with a plaque-forming unit



Fig 3|Predicted influenza-human protein-protein interaction network. Network composed of 108 interactions involving 6 influenza viral proteins and 41 human proteins. Twelve of 41 human proteins have been identified to be host factors for influenza virus replication (from literature and this work). The CHUK-NS1 interaction was already reported in the literature.

assay. Using this criteria, 10 new cellular interactors of NS1 were found to directly modulate the replication of influenza A virus three were functioning as restriction host factors and seven as essential host factors (Figs 2A,C,D). None of the tested siRNAs affected cell survival (supplementary Fig S2 online). Therefore, the method generated a data set that is strongly enriched in modulators of replication—43.5 versus 0.05–1.2% for siRNA pangenomic screens.

Proteome-wide prediction of virus-host interactions

Having validated the method, predictions were extended to other proteins of the same virus to construct a predicted influenza– human interactome. Structural information was available for 13 structures corresponding to nine viral proteins. Following the process described above for NS1, 108 interactions were predicted connecting 41 cellular proteins to six viral proteins (Fig 3, supplementary Tables S3 and S4 online). The low overlap of our set of interactions with literature reflects the incompleteness of previously reported influenza virus–host protein interaction data (supplementary information online). In addition, structural and interactomic data availability restricts the list of predictable interactors, as illustrated for NS1 in supplementary information online. Some human proteins are predicted to interact with several viral proteins. Multi-targeting of specific host factors by different proteins of the same virus has been observed previously, highlighting the importance of such proteins for the virus [24]. Interestingly, viral proteins sharing common interactors are those displaying a certain degree of structural homology (supplementary Table S5 online). Finally, an up-to-date influenza–human interactome combining literature and predicted data are presented in Fig 4.

METHODS

Retrieval and annotation of viral and human structures. All sequences from viral and human protein structures stored at the PDB were retrieved in a Fasta format and blasted to be annotated with a GenBank identification for viral proteins and Ensembl protein identification for human proteins as it is the case in the VirHostNet database. A total of 4,933 PDB sequences (viral GenBank identification) and 22,052 PDB sequences (human Ensembl protein identification) have been retrieved and assigned to 440 distinct viral proteins and 3,163 distinct human proteins. For the full-length influenza A virus NS1 protein, the chain A of the 3F5T PDB structure has been used.

High-quality protein–protein interaction data from VirHostNet. Protein–protein interaction data were retrieved from the VirHostNet



Fig 4|Influenza-human interactome combining predicted and literature data. Network composed of 515 interactions involving 10 influenza viral proteins and 364 human proteins. One-hundred and six interactions were obtained from predictions, 407 from literature and 2 from both. Forty-seven of 364 human proteins have been identified to be host factors directly controlling viral replication (literature and this work).

database [2]. A high-quality data set of human protein–protein interactions were defined by selecting interactions identified by two different methods or in two independent papers. Indeed, the human interactome is an integration of data from several databases, and might be polluted by low-confidence interactions [25]. A high-quality data set of virus–human protein– protein interactions were constructed with exclusively manually curated interactions. To experimentally validate the predicted interactions, an additional level of stringency was added by focusing on interactions described by classical laboratory technologies. Altogether, the predictions are on the basis of 16,187 human protein–protein interactions involving 6,378 distinct human proteins and 2,539 virus–host protein–protein interactions involving 434 viral proteins and 1,395 human proteins (supplementary Tables S6 and S7 online).

Score definition. All putative interactors were individually scored according to several criteria. First, an interactor is more confident when predicted by several independent NS1 homologues. Interactors independently predicted by both viral and

human homologues are favored. The prediction from viral homologues only is also better considered than the prediction from human homologues only. Finally, as a correlation was observed between the score and the degree of a putative interactor (that is, the number of partners in the protein interaction data set), the score was weighted by the degree. Altogether, the score formula is as follows: score (putative interactor = $[(n_H^2/K_H) + (n_V^2/k_V)] \times a$, with: (i) n_H and n_V number of distinct human interactors and viral interactors, respectively. (ii) k_h and k_V degree of the protein in the human protein–protein interaction data set, respectively. (iii) a = 3 for a prediction from both human and virus–human data sets, a = 2 from only virus–human data set, 1 else. (iv) H: human, V: virus.

Cells and virus. A549 human lung epithelial cell line and Madin-Darby canine kidney (MDCK) cell line were grown in DMEM media supplemented with 100 U ml⁻¹ penicillin/streptomycin and 10% fetal calf serum at 37 °C and 5% CO₂. The epidemic A/H1N1/New Caledonia/2006 strain was propagated in MDCK

cells in DMEM supplemented with $1\,\mu g\,ml^{-1}$ modified trypsin TPCK in absence of fetal calf serum. Virus stocks were titrated by standard plaque assay on MDCK cells using an agar overlay medium.

GST pull down. Cellular proteins (corresponding to the 32 cDNAs available in orfeotech ORFeome v3.1 [26], description of the hORFeome v3.1 in supplementary information online) were amplified from the MGC cDNA plasmid collection using KOD polymerase (Toyobo), cloned by recombinational cloning into pDONR207 (Life technologies) and transferred to pDESTmyc or pClneo3xFlag (kind gift of Y. Jacob). H5N1 NS1 was transferred into pDEST27 (GST fusion in N-term). A total of 4.10⁵ HEK293T cells were then co-transfected (6 µl JetPei, Polyplus) with 1.5 µg of each pair of plasmid. Two days after transfection, cells were harvested and lysed (0.5% NP-40, 20 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM EDTA and Roche complete protease inhibitor cocktail). Soluble protein complexes were purified by incubating 300 µg of cleared cell lysate with 40 µl glutathione sepharose 4B beads (GE Healthcare). A total of 50 µg of cleared cell lysate was separated on SDS-PAGE and transferred to nitrocellulose membrane. GST-tagged viral proteins and Myc or 3XFlag-tagged cellular proteins were detected using standard immunoblotting techniques with anti-GST (Covance), anti-Myc or anti-Flag (Sigma) monoclonal antibodies.

Yeast two-hybrid. Pairwise Y2H interactions were analysed by yeast mating, using Y187 and AH109 yeast strains (Clontech). Prey and bait vectors were transformed into Y187 and AH109. Bait and prey strains were mated in an all-against-all array and plated on a selective medium lacking histidine and supplemented with increasing concentrations of 3-amino-triazole to test the interaction-dependent transactivation of *HIS3* reporter gene. Some of the preys are membrane proteins but reported in the literature to give positive results in Y2H (for example, BCL2L11, BAD) or proteins whose signal peptide is likely to be masked (supplementary information online). Therefore, we did not exclude them from the Y2H.

siRNA screening. Five picomoles of each siRNA (stealth select RNAi, Invitrogen) was arrayed in 96 plates in 10µl of OptiMEM (GibCo). Ten microliters per well of a mix lipofectamine RNAiMAX (Invitrogen)-OptiMEM (0.2µl of lipofectamine RNAiMAX in 10µl of OptiMEM) was added. After 20min room temperature-incubation, siRNA-lipofectamine mixes were added to 30×10^3 A549 suspension cells. Cells were incubated for 48 h at 37° C and 5% CO₂ before influenza virus infection at multiplicity of infection 0.5. Forty-eight hours after infection, supernatants were titrated.

Virus infection. siRNA-transfected cells were washed twice with Dulbecco's phosphate-buffered saline $1 \times$ and infected with H1N1 influenza strain at multiplicity of infection 0.5 in Dulbecco's modified Eagle medium supplemented with 0.2 µg ml⁻¹ trypsin TPCK (infection medium). After 60 min at 37 °C, the inoculum was discarded and cells were washed again and incubated for 48 h in infection medium at 37 °C and 5% CO₂. **Titre measure by neuraminidase activity.** Influenza virus neuraminidase is able to cleave the methyl-umbelliferyl-*N*-acetylneuraminic acid (4-MUNANA, Sigma) modifying its emission wavelength in a dose-dependent manner. In 96-black plate, 25μ l infection supernatants were diluted in 25μ l Dulbecco's phosphate-buffered saline $1 \times$ containing calcium

and magnesium and $50\,\mu$ l of $20\,\mu$ M 4-MUNANA. After 1 h incubation at $37\,^{\circ}$ C, $100\,\mu$ l of glycine 0.1 M 25% ethanol pH 10.7 was added. Measures were done with TECAN infinite M1000 instrument at 365 nm excitation and 450 nm emission wavelengths.

Plaque-forming unit assay. To quantify infectious virus particles in infected cell culture supernatants, 300,000 MDCK cells were seeded in 6-well plates. Three days later, cells were washed twice with DMEM supplemented with 100 Uml^{-1} penicillin/ streptomycin. Dilutions of infected cell culture supernatants were dispensed on the cells. After 1 h 30 min incubation at 37 °C and 5% CO₂ cells were washed again and covered with overlay medium containing MEM 1 × , 1% agarose (Lonza) and modified trypsin TPCK 1 µg ml⁻¹. Plates were incubated upside down at 37 °C and 5% CO₂ up to 72 h. Cells were then fixed with formol 10% and coloured with crystal violet 0.3% in methanol 20% and lysis plaques were counted.

Functional annotation. DAVID functional annotation chart tool was used to perform enrichment analysis, in which an annotation term was considered as significant with a Benjamini–Hochberg corrected *P*-value smaller than 0.05 [27].

Supplementary information is available at EMBO *reports* online (http://www.emboreports.org).

ACKNOWLEDGEMENTS

The authors thank Liisa Holm for her technical help on Dali. This work was supported by the European Union's 7th Program (FP7/2007–2013) under grant agreement no. 267429 (SysPatho) and European Community's 7th Framework Program (FP7/2007–2013) and Cosmetics Europe under grant agreement no.266753 (SCR&Tox).

Author contributions: B.d.C conceived the method; B.d.C. and V.L. supervised the experiments; B.d.C., L.M.-S. and V.N. designed and performed bioinformatic studies; A.A.-G., T.C. and B.d.C. designed and performed biological studies; P.A., A.A.-G., B.d.C., V.L. and L.M.-S. analysed the data; B.d.C., V.L. and L.M.-S. wrote the manuscript.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

- 1. Chatr-aryamontri A *et al* (2009) VirusMINT: a viral protein interaction database. *Nucleic Acids Res* **37:** D669–D673
- Navratil V, De Chassey B, Meyniel L, Delmotte S, Gautier C, Andre P, Lotteau V, Rabourdin-Combe C (2009) VirHostNet: a knowledge base for the management and the analysis of proteome-wide virus-host interaction networks. *Nucleic Acids Res* 37: D661–D668
- 3. de Chassey B *et al* (2008) Hepatitis C virus infection protein network. *Mol Syst Biol* **4:** 230
- 4. Jager S *et al* (2012) Global landscape of HIV-human protein complexes. *Nature* **481:** 365–370
- 5. Braun P *et al* (2009) An experimentally derived confidence score for binary protein-protein interactions. *Nat Methods* **6:** 91–97
- Jessulat M et al (2011) Recent advances in protein-protein interaction prediction: experimental and computational methods. Expert Opin Drug Discov 6: 921–935
- Mukhopadhyay A, Maulik U, Bandyopadhyay S (2012) A novel biclustering approach to association rule mining for predicting HIV-1-human protein interactions. *PLoS One* 7: e32289
- 8. Aloy P, Russell RB (2002) Interrogating protein interaction networks through structural biology. *Proc Natl Acad Sci USA* **99:** 5896–5901
- 9. Davis FP, Braberg H, Shen MY, Pieper U, Sali A, Madhusudhan MS (2006) Protein complex compositions predicted by structural similarity. *Nucleic Acids Res* **34:** 2943–2952

- Tuncbag N, Gursoy A, Nussinov R, Keskin O (2011) Predicting proteinprotein interactions on a proteome scale by matching evolutionary and structural similarities at interfaces using PRISM. *Nat Protoc* 6: 1341–1354
- Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 247: 536–540
- 12. Bornholdt ZA, Prasad BV (2008) X-ray structure of NS1 from a highly pathogenic H5N1 influenza virus. *Nature* **456:** 985–988
- 13. Holm L, Park J (2000) DaliLite workbench for protein structure comparison. *Bioinformatics* **16:** 566–567
- Holm L, Kaariainen S, Wilton C, Plewczynski D (2006) Using Dali for structural comparison of proteins. *Curr Protoc Bioinformatics* Chapter 5 Unit 5.5
- Schultz-Cherry S, Dybdahl-Sissoko N, Neumann G, Kawaoka Y, Hinshaw VS (2001) Influenza virus ns1 protein induces apoptosis in cultured cells. J Virol 75: 7875–7881
- Zhirnov OP, Konakova TE, Wolff T, Klenk HD (2002) NS1 protein of influenza A virus down-regulates apoptosis. J Virol 76: 1617–1625
- Ehrhardt C, Wolff T, Pleschka S, Planz O, Beermann W, Bode JG, Schmolke M, Ludwig S (2007) Influenza A virus NS1 protein activates the PI3K/Akt pathway to mediate antiapoptotic signaling responses. *J Virol* 81: 3058–3067
- Lam WY, Tang JW, Yeung AC, Chiu LC, Sung JJ, Chan PK (2008) Avian influenza virus A/HK/483/97(H5N1) NS1 protein induces apoptosis in human airway epithelial cells. J Virol 82: 2741–2751

- Lu Y, Wambach M, Katze MG, Krug RM (1995) Binding of the influenza virus NS1 protein to double-stranded RNA inhibits the activation of the protein kinase that phosphorylates the elF-2 translation initiation factor. *Virology* 214: 222–228
- Talon J, Horvath CM, Polley R, Basler CF, Muster T, Palese P, Garcia-Sastre A (2000) Activation of interferon regulatory factor 3 is inhibited by the influenza A virus NS1 protein. J Virol 74: 7989–7996
- Pichlmair Á, Schulz O, Tan CP, Naslund TI, Liljestrom P, Weber F, Reis e Sousa C (2006) RIG-I-mediated antiviral responses to single-stranded RNA bearing 5'-phosphates. *Science* **314**: 997–1001
- Gack MU et al (2009) Influenza A virus NS1 targets the ubiquitin ligase TRIM25 to evade recognition by the host viral RNA sensor RIG-I. Cell Host Microbe 5: 439–449
- 23. Pleschka S (2008) RNA viruses and the mitogenic Raf/MEK/ERK signal transduction cascade. *Biol Chem* **389**: 1273–1282
- 24. Meyniel-Schicklin L, de Chassey B, Andre P, Lotteau V (2012) Viruses and interactomes in translation. *Mol Cell Proteomics* **11**: 014738
- 25. Cusick ME *et al* (2009) Literature-curated protein interaction datasets. *Nat Methods* **6:** 39–46
- Lamesch P et al (2007) hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes. *Genomics* 89: 307–315
- 27. Huang da W, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37:** 1–13