Good Practice in Traditional Chinese Medicine Research in the Post-genomic Era

GP-TCM

223154

D4.16

Handbook for using functional genomics techniques in *in-vitro* CHM research

Document description	
Name of document	Handbook for using functional genomics techniques in *in vitro* CHM research
Abstract	A handbook presents the state of the art of the use of functional genomic techniques in CHM research. It represents the culmination of the numerous discussions and deliberations of the WP4 working party of the FP7-funded European consortium, GP-TCM. It provides future perspectives for the use of omics and *in-silico* approaches in the modernisation and standardisation of TCM, and provides - for the first time - guidelines for good practice in the application of these methods in CHM research.
Document identifier	D4.16
Document class	Deliverable
Version	1.0
Author(s)	Alessandro Buriani (UP), Peter Hylands (KCL), David Barlow (KCL) and all WP4 members
Date of creation	18/03/2012
Date of last modification	28/03/2012
Status	Final
Destination	European Commission
WP number	WP4

TABLE OF CONTENTS

1 HANDBOOK FOR USING FUNCTIONAL GENOMICS TECHNIQUES IN IN-VITRO CHM RESEARCH

1.1 Definition Of Terms

Functional Genomics: In the following handbook, "functional genomics", otherwise abbreviated as omics, is used as an umbrella term for genomic, proteomic and metabonomic studies, wherein the plurality of changes in gene expression, protein expression and metabolite formation are investigated, at the cellular, multi-cellular, or organism level, as a function of various internal and/or external perturbing factors.
In-silico: *In-silico* studies are here defined as those computational studies that involve virtual screening and/or cheminformatics but more broadly may also include those involving bioinformatics as applied in the various types of omics studies.

1.2 Preface

Identification of molecular mechanisms and targets is a critical step in the validation of a biological effect. When using phytocomplexes as in Chinese herbal medicine (CHM) research, this is often hampered by the complexity of the molecular mixtures, with many different molecules contributing to the overall effect, either positively or negatively. With the advent of information-rich techniques such as genomics, proteomics and transcriptomics as well as various profiling approaches, including metabolomics (a non-targeted analytical approach, usually concentrating on molecules of low molecular weight) and metabonomics (similar studies but involving studying the effects of *perturbation* of a system), it has recently become possible to examine simultaneous molecular effects of mixtures of chemical agents and, with the help of bioinformatics, to look at such effects with a global view on the biological system. Factorial analytical models can decode the large quantity of raw information derived from these omic techniques, sometimes allowing correlation of the multiple components of phytocomplexes with their biological effects. At the same time, high-throughput, information-rich assays can be used to fingerprint herbs and botanical extracts. Such applications of information-rich approaches make the use of omic techniques particularly appropriate for addressing many of the problems encountered in traditional Chinese medicine (TCM) research that have hampered its acceptance in the Western biomedical mainstream and its integration with more orthodox Western medical practice. Today, a holistic systems biology approach provides a new perspective in pharmacological science which goes beyond target specificity and single molecule pharmacology, and embraces the entire equilibrium of a biological system undergoing simultaneous perturbations on primary and secondary multiple molecular targets. This approach is even more relevant when using multi-chemical mixtures, as in CHM and other TCM formulae which comprise mixtures of mixtures.

To date, most investigations into the applications of CHM have been based on analyses of clinical data, as well as identification and testing of chemical constituents of the herbs. Over recent years, however, there have been increasing efforts concerned with the application of computational methods to screening for bioactive compounds in Chinese herbs and identifying their targets, and on the role that informatics can play in discovering links between molecular pharmacology, on the one hand, and the insights and language of TCM on the other.

The Handbook presented here represents the culmination of the numerous discussions and deliberations of the WP4 working party of the FP7-funded European consortium for GP-TCM (Good Practice in Traditional Chinese Medicine Research in the Post-genomic Era). It provides future perspectives for the use of omics and *in-silico* approaches in the modernisation and standardisation of TCM, and provides for the first time guidelines for good practice in the application of these methods in CHM research.

1.3 Disclaimer

It is important to note here that the guidelines presented below merely represent the consensus views of the WP4 working party of the GP-TCM consortium (Uzuner et al, 2012); they are not defined in any analytical sense, nor are they supported by scientific case studies. The working party nonetheless contend that the use of non-reductive technologies is highly appropriate to begin to understand the complexities of the use of herbal materials at the molecular level.

2 INTRODUCTION

Regarding the application of omics in TCM research, a questionnaire was circulated among a group of GP-TCM researchers "D4.14 – Discussion group on use of functional genomic techniques for in vitro CHM research". This questionnaire addressed issues related to the use of omic techniques in TCM research and most researchers agreed on several pros and cons. The most relevant conclusions are:

- Quality control of the test material at all stages of preparation and production and its standardisation still remain the main barriers to meaningful research, but omic techniques can, and increasingly do, contribute significantly to this important aspect.
- Thanks to its wider view of biological systems and closer view of simultaneous multiple effects exerted by phytocomplexes, systems biology can contribute to increased participation of CHM in the scientific mainstream.
- Omic technologies are efficient at providing information on whole collections of molecules. These can then either be studied further using a reductionist approach to establish the properties and function of each or to try to understand and validate the relationship between them and their functions.

For these reasons, one single methodology is not considered sufficient to investigate mechanisms of action of CHM. A well-consolidated pipeline should be used comprising *in silico* evaluation, *in-vitro* and *in-vivo* validation through a combination of classical biochemical signalling work, conventional molecular biology, omics technology and bioinformatics.

3 GENERAL SUGGESTIONS FOR USING OMIC TECHNIQUES IN CHM RESEARCH

Omic methodologies can be usefully applied to different phases of CHM research starting from standardisation and quality control of herbal formulae, characterisation of target-mediated and downstream effects, as well as identification of molecular mechanisms to predict side effects and interactions with other drugs. Thanks to their potential to unveil the interconnected pharmacological networks induced by complex herbal preparations, omic techniques can thus be considered powerful tools to address many open questions in CHM research.

While all the omic techniques are slowly but firmly pushing their way further in CHM research, metabonomics seems to be rapidly gaining ground with respect to the others. This is probably due to the simplicity of the experimental design and its affordability, which allows direct and detailed analysis of large numbers of biological samples which, like urine, can easily be obtained. Metabonomics provides the possibility of examining complete metabolic pathways and their intermingled interactions in just one snapshot, taking a whole picture of the downstream outcomes of any biological perturbation. This would appear as the ultimate systems biology phenotyping and is particularly fit for studying TCM, with its holistic view of biological effects. Accordingly, not only metabonomics is used to study the action of Chinese formulae, but is being increasingly used to successfully characterise TCM syndromes.

When using omics in CHM studies using experimental models of disease, at least two main issues need to be addressed.

The first aspect is related to the intrinsic problem of replicating the patterns of human disease in animal or cellular models. Both *in-vitro* and *in-vivo* types of models are rarely close representations of the clinical scenario and are often not widely accepted. Indeed some animal or disease models are not even clearly characterised or validated. Therefore prior validation and standardisation of animal models where omics could be applied to the study of TCM is required, as has been the case in Western medicine research.

Omics methods are powerful but since they involve studying the basic components of the systems, they are also liable to variability. A particular source of variability however that has not been generally considered in most publications on CHM in experimental models of disease is the generalised use of non-standardised research materials (in terms of the constituent herbs and herbal preparations), which itself promotes variability. This lack of rigour significantly reduces the scientific value and impact of these studies. This implies that before using omic technologies, it is necessary to have robust control of plant mixture preparation (batch to batch variability), as well as the experimental model: cell cultures and/or animal system (organism variability as well as technical procedures). In addition, pharmacokinetic (absorption, distribution, metabolism and excretion) profiling of a given CHM may identify *in-vitro* and *in-vivo* bioavailable drug-like components and reveal determining factors for availability, dynamics and individual variations of bioactive components.

Given that omic technologies may help elucidate the mechanism of action of a given CHM treatment, it is suggested that these studies have to be applied on those CHM treatments the efficacies of which have been previously demonstrated. Since most relevant pieces of evidence on efficacy come from clinical trials, it is thus advisable to start applying omic technologies to animal models of diseases for which efficacy has been proven in the clinic and/or to well-characterized clinical syndromes and disease entities (be they "western" or "chinese") without prior validation of animal models. This would render a number of TCMs relatively easy to deal with. In this context the scheme in Fig. 1 is proposed. Additionally the following workflow for applying omics in experimental *in-vitro* and *in-vivo* models is suggested:

(1) to use a TCM drug proven to be efficacious in an appropriate, well characterized, clinically relevant, experimental model,
(2) to assess variability in TCM composition and select a uniform, appropriately defined batch,
(3) to assess variability in the organism population (cell cultures or animals) by carrying out omic pilot studies, including metabolic profiling,
(4) to take into consideration the known levels of variation in the changes observed (whether at the transcript, protein or metabolic levels),
(5) to define precisely the experimental groups as well as their size in terms of number of cell types or animals to enable statistical analysis. It will also be desirable to choose a homogeneous experimental population as far as possible,
(6) to check that the effects observed in cell cultures or animals after TCM use, especially if chronic administration is necessary, are due to the treatment and not to some other variable (e.g., cell passage number, aging or body weight changes). As much data must be collected as possible, and
(7) to perform the appropriate omic technique and analysis according to the available guidelines on omics standardisation in the literature, such as MIAME (minimum information about a microarray experiment) for transcriptomics (Brazma et al., 2001), MIAPE (minimum information about a proteomic experiment) for proteomics (Taylor et al., 2007), and MSI (the metabolic standards initiative; website: http://msi-workgroups.sourceforge.net) for metabolomics.

3.1 In-silico tools for proteomics, genomics and metabonomics data visualisation and analysis

Using omic techniques much experimental data can be obtained that needs to be manipulated both in terms of multichemical identification and pathophysiological correlations. With the help of bioinformatics, specific software and databases have been developed. This systems biology approach thus integrates powerful information-rich technologies, computational tools and knowledge bases, making it possible to establish links between molecular patterns, biological functions and a wide range of human diseases and pharmacological interventions. In particular,

high-throughput omic techniques require several bioinformatic and knowledge-assembly tools for data processing, analysis, integration and interpretation using a top-down systems biology approach. When using omic techniques, raw data are produced mostly from microarrays and mass spectrometry or NMR spectroscopy, and subsequent to data processing, which often needs multivariate analysis, sets of qualitative and quantitative data are obtained indicating patterns of induced molecular perturbations. *In-silico* tools can be used to assist in molecular identification, and databases as well as specific search software are available to assist in this task. These include the METLIN and MassBank databases for metabolomics ([Sana et al., 2008], and [Horai et al., 2010]), and human protein and peptide databases for proteomics ([Choi et al., 2008], [Klimek et al., 2008], [Ding et al., 2008], [Wang et al., 2011]).

Following analysis, the affected molecular networks can be identified and the observed perturbations correlated to a given pharmacological and/or toxicological effect. This last correlation step is of utmost importance in order to exploit the full potential of information-rich data. Several databases have been made available that can be used for guided analysis, while others can be applied to untargeted and genome-wide or metabolome-wide association studies ([Smith and Newton-Cheh, 2009] and [Chadeau-Hyam et al., 2010]). These databases allow data mining, correlation searches, and modelling of biochemical pathways. Molecular databases can thus be used to identify known metabolites, or highlight unknown ones, or to determine the biological function of a specific molecule.

The catalogue of databases and software developed for genome-scale metabolic reconstructions is ever increasing, and many of these utilities are freely available for academic use.

The following section provides just a few selected examples.

a) The *Human Metabolome Database* (HMDB) is considered the most complete bioinformatics and chemoinformatics medical information database (Kouskoumvekaki and Panagiotou, 2011) with more than 6800 metabolite entries; it is fully searchable with tools for viewing, sorting metabolites and pathways, and provides customised, clickable metabolic maps, as well as disease information (Wishart et al., 2009).
b) BiGG, *Biochemical Genetic and Genomic knowledge base*, is a metabolic reconstruction of human metabolism using both genetics and literature-based data to assess whether a reaction is present, providing confidence level data and Boolean relationships between genes, proteins and reactions (Schellenberger et al., 2010).
c) The *Small Molecule Pathways Genomics Portal* provides a database of human, mouse, and rat genomic data with basic analytical visualisation tools (Shinde et al., 2010).
d) Examples of proteomic databases include GPDE, *Griss Proteomics Database Engine*, PRIDE, *Proteomics IDEntifications database*, and *PeptideAtlas* (Griss et al., 2011).
e) Some databases are dedicated to specific fields of study. For example, T3TB, the Toxin and Toxin-Target Database, provides descriptions, mechanisms of action, and information on toxins and toxin-targets (Lim et al., 2010).
f) CEBS, Chemical Effects in Biological Systems, is an integrated public repository focused on toxicogenomics, integrating data describing histopathological and biological measures with microarray and proteomics data (Waters et al., 2008).

Even though the introduction of omic techniques in CHM research is relatively recent, meaningful examples of the application of a systems biology approach can be found in the current literature, and the results are suggestive of a great potential for application of omics and related bioinformatic techniques in the study of both Chinese syndromes and phytocomplexes from traditional Chinese medicine.
For the benefit of the interested reader examples are provided in *Barlow et al, 2012*, and further discussions on the application of omics methodologies in TCM research are presented in Ouedraogo et al (2012), Pelkonen et al (2012), Jia et al (2012), and Buriani et al (2012).
In terms of guidelines production for the researcher in the field, one of the next steps that could be envisioned from this handbook, is the development of more specific handbooks focusing on guidelines in specific areas of TCM research, along the lines of handbooks for conventional pharmaceuticals, employing also omics techniques and systems biology approach.

4 CONCLUSION

Analysis of omics and biological function data represents the common pathway to functional genomics. Much experimental data generated by high-throughput techniques are available through various public repositories. Knowledge about transcriptional regulation, molecular interaction networks and metabolic pathways is rapidly expanding thanks to the innovative system biology approach, information rich omics technology and computer assisted analysis.
In this rapidly evolving context the experimental study of multiple target, multichemical pharmacologic effects with phytocomplexes as in CHM, can greatly benefit from this new holistic scientific vision.

5 REFERENCES

Barlow DJ, Buriani A, Ehrman T, Bosisio E, Eberini I, Hylands PJ. In-silico studies in Chinese herbal medicines' research: Evaluation of in-silico methodologies and phytochemical data sources, and a review of research to date. J Ethnopharmacol. 2012 Feb 2. [Epub ahead of print]

Buriani A, Garcia-Bermejo ML, Bosisio E, Xu Q, Li H, Dong X, Simmonds MS, Carrara M, Tejedor N, Lucio-Cazana J, Hylands PJ. Omic techniques in systems biology approaches to traditional Chinese medicine research: Present and future. J Ethnopharmacol. 2012 Feb 8. [Epub ahead of print]

Chadeau-Hyam, M., Ebbels, T.M.D., Brown, I.J., Chan, Q., Stamler, J., Huang, C.C., Daviglus, M.L., Ueshima, H., Zhao, L., Holmes, E., Nicholson, J.K., Elliott, P., De Iorio, M., 2010. Metabolic Profiling And The Metabolome-Wide Association Study: Significance Level For Biomarker Identification. Journal of Proteome Research 9, 4620–4627.

Choi, H., Ghosh D., Nesvizhskii, A.I., 2008. Statistical Validation of Peptide Identifications in Large-Scale Proteomics Using the Target-Decoy Database Search Strategy and Flexible Mixture Modeling. Journal of Proteome Research 7, 286–292.

Clive, D., Greiner, R., Nazyrova, A., Shaykhutdinov, R., Li, L., Vogel, H.J., Forsythe, I., 2009. HMDB: a knowledgebase for the human metabolome. Nucleic Acids Research 37, D603–D610.

Ding, Y., Choi, H., Nesvizhskii, A.I., 2008. Adaptive Discriminant Function Analysis and Reranking of MS/MS Database Search Results for Improved Peptide Identification in Shotgun Proteomics. Journal of Proteome Research 7, 4878–4889.

Griss, J., Haudek-Prinz, V., Gerner, C., 2011. GPDE: A biological proteomic database for biomarker discovery and evaluation. Proteomics 11, 1000–1004.

Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., Ojima, Y., Tanaka, K., Tanaka, S., Aoshima, K., Oda, Y., Kakazu, Y., Kusano, M., Tohge, T., Matsuda, F., Sawada, Y., Hirai, M., Nakanishi, Y., Ikeda, H., Akimoto, K. , Maoka, N., Takahashi, H., Ara, T., Sakurai, N., Suzuki, H., Shibata, D., Neumann, S., Iida, T., Tanaka, K., Funatsu, K., Matsuura, F., Soga, T., Taguchi, R., Saito, K.i, Nishioka, T., 2010. MassBank: a public repository for sharing mass spectral data for life sciences. Journal of Mass Spectrometry 45, 703–714.

Jia J, Yu Y, Deng JH, Robinson N, Bovey M, Cui YH, Liu HR, Ding W, Wu HG, Wang XM. A review of Omics research in acupuncture: The relevance and future prospects for understanding the nature of meridians and acupoints. J Ethnopharmacol. 2012 Feb 1. [Epub ahead of print]

Klimek J., Eddes, J.S., Hohmann, L., Jackson, J., Peterson, A., Letarte, S., Gafken, P.R., Katz, J.E., Mallick, P., Lee, H., Schmidt, A., Ossola, R., Eng, J.K., Aebersold, R., Martin, D.B., 2008The Standard Protein Mix Database: A Diverse Dataset to Assist in the Production of

Improved Peptide and Protein Identification Software Tools. Journal of Proteome Research 7, 96–103.

Kouskoumvekaki, I., Panagiotou, G., 2010. Navigating the HumanMetabolome for Biomarker Identification and Design of Pharmaceutical Molecules. Journal of Biomedicine and Biotechnology 2011, Article ID 525497, 19 pages.

Lim, E., Pon, A., Djoumbou, Y., Knox, C., Shrivastava, S., Guo, A.C., Neveu, V., Wishart, D.S., 2010. T3DB: a comprehensively annotated database of common toxins and their targets. Nucleic Acids Research 38, D781–D786.

Ouedraogo M, Baudoux T, Stévigny C, Nortier J, Colet JM, Efferth T, Qu F, Zhou J, Chan K, Shaw D, Pelkonen O, Duez P. Review of current and "omics" methods for assessing the toxicity (genotoxicity, teratogenicity and nephrotoxicity) of herbal medicines and mushrooms. J Ethnopharmacol. 2012 Feb 22. [Epub ahead of print]

Pelkonen O, Pasanen M, Lindon JC, Chan K, Zhao L, Deal G, Xu Q, Fan TP. Omics and its potential impact on R&D and regulation of complex herbal products. J Ethnopharmacol. 2012 Feb 1. [Epub ahead of print]

Sana, T.R., Roark, J.C., Li, X., Waddell, K., Fischer, S.M., 2008. Molecular Formula and METLIN Personal Metabolite Database Matching Applied to the Identification of Compounds Generated by LC/TOF-MS. Journal of Biomolecular Techniques 19, 258–266.

Schellenberger, J., Park, J.O., Conrad, T.M., Palsson, B.Ø., 2010. BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. BMC Bioinformatics 11, 213.

Shinde, K., Phatak, M., Johannes, F. M., Chen, J., Li, Q., Vineet. J.K., Hu, Z.,Ghosh, K., Meller, J., Medvedovic, M., 2010. Genomics Portals: integrative web-platform for mining genomics data. BMC Genomics 11, 27.

Smith, J.G., Newton-Cheh, C., 2009. Genome-wide association study in humans. Methods in Molecular Biology 573, 231-258.

Uzuner H, Bauer R, Fan TP, Guo DA, Dias A, El-Nezami H, Efferth T, Williamson EM, Heinrich M, Robinson N, Hylands PJ, Hendry BM, Cheng YC, Xu Q. Traditional Chinese medicine research in the post-genomic era: Good practice, priorities, challenges and opportunities. J Ethnopharmacol. 2012 Feb 23. [Epub ahead of print]

Wishart, D.S., Knox, C., Guo, A.C., Eisner, R., Young, N., Gautam, B., Hau, D.D., Psychogios, N., Dong, E., Bouatra, S., Mandal, R., Sinelnikov, I., Xia, J., Jia, L., Cruz, J.A., Lim, E., Sobsey, C.A., Shrivastava, S., Huang, P., Liu, P., Fang, L., Peng, J., Fradette, R., Cheng, D., Tzur, D., Clements, M., Lewis, A., De Souza, A., Zuniga, A., Dawe, M., Xiong, Y., ang, X., Slebos, R.J.C., Wang, D., Halvey, P.J., Tabb, D.L., Liebler, D.C., Zhang B., 2011 Protein identification using customized protein sequence databases derived from RNA-Seq data. Journal of Proteome Research, in press.

Waters, M., Stasiewicz, S., Merrick, B. A., Tomer, K., Bushel, P., Paules, R., Stegman, N., Nehls, G., Yost, K.J., Johnson, C.H., Gustafson, S.F., Xirasagar, S., Xiao, N., Huang, C.-C., Boyer, P., Chan, D.D., Pan, Q., Gong, H., Taylor, J., Choi, D., Rashid, A., Ahmed, A., Howle, R., Selkirk, J., Tennant, R., Fostel, J., 2008.CEBS—Chemical Effects in Biological Systems: a public data repository integrating study design and toxicity data with microarray and proteomics data. Nucleic Acids Research 36, D892–D900.