Gene 572 (2015) 101–107

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Single amino acid changes in naked mole rat may reveal new anti-cancer mechanisms in mammals

Zhiyuan Yang [a], Yan Zhang [b,*], Luonan Chen [a,c,d,**]

[a] Key Laboratory of Systems Biology, Innovation Center for Cell Signaling Network, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China
[b] Key Laboratory of Nutrition and Metabolism, Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China
[c] School of Life Science and Technology, Shanghai Tech University, Shanghai 201210, China
[d] Collaborative Research Centre for Innovative Mathematical Modelling, Institute of Industrial Science, The University of Tokyo, Tokyo 153-8505, Japan

ARTICLE INFO

Article history:
Received 26 August 2014
Received in revised form 2 April 2015
Accepted 1 July 2015
Available online 3 July 2015

Keywords:
Cancer
Naked mole rat
Single amino acid change
Sequence
Pathway

ABSTRACT

Background: The naked mole rat (NMR, *Heterocephalus glaber*) is a long-lived rodent model with multiple extraordinary traits. They live underground and have been known to live for up to thirty years, much longer than similar-sized mice. Moreover, congenital cancer or experimentally induced cancer genesis could not been observed in this rodent so far. Such unique biochemical and physiological characteristics lead them to become a popular model for cancer research.

Results: In this paper, a genome-wide comparative analysis was conducted based on the genomes of NMR and several other mammals. First, all the annotated proteins of NMR were searched against 11 selected mammalian genomes to verify their occurrence in these organisms. Among them, 66 NMR genes were not detected in other 11 mammals, almost all of which present alkalinity isoelectric points. In contrast, a total of 89 genes that are present in all of the 11 organisms could not be found in NMR genome. Among them, 3 genes are known to be related to cancer development. Finally, we identified NMR-specific single amino acid change (SAAC) events for the proteins that are present in both NMR and other mammals. KEGG pathway database was also used to investigate the metabolic processes in which these SAAC proteins may be involved. These genes were significantly enriched in two known cancer pathways, "Pathways in cancer" and "Pancreatic cancer". In the "Pancreatic cancer" pathway, 3 out of 6 paths leading to DNA duplication appeared to be affected by direct connection to the SAAC genes in NMR. In addition, a significant number of other SAAC genes enriched in several cancer-related pathways have been known to be associated with a variety of cancers, implying that many of them may be also related to tumor genesis in mammals.

Conclusions: Overall, our results not only can be used to find possible genes involved in physiological mechanism of NMR but also provide new clues for the anti-cancer mechanism of NMR.

1. Introduction

The naked mole rats (NMRs) are peculiar underground poikilotherms native to Ethiopia and Kenya. They are hairless rodents and live an organized social life like bees and ants (Holmes et al., 2007). NMRs possess several specific features when compared to other mammals, making them a hot spot for life science research recently.

NMRs are known to live in a lifetime of up to thirty years, much longer than similar-sized mice (Ke et al., 2014). Their body functions only have negligible decrement with the increasing of the age, indicating that NMRs do indeed have anti-aging properties (Grimes et al., 2014). Moreover, either congenital cancer or experimentally induced cancer genesis has not been observed for NMR. Previous studies have shown that carcinogenesis could not be achieved in NMR by the transfection of the oncogene *Hras1*, while mice of control group had induced cancer easily (Liang et al., 2010). Furthermore, NMR is insensitive to skin pain. It has been shown that two neuro-related genes (calcitonin gene-related peptide and neurotransmitter substance P) are specifically not expressed in the NMR's skin, blocking the pain signals transmitting from the skin nerves to the central nervous system, thus can hardly

Abbreviations: NMR, naked mole rat; SAAC, single amino acid change; PPI, protein–protein interaction; pI, isoelectric point; FDR, false discovery rate; AGCOH, Atlas of Genetics and Cytogenetics in Oncology and Haematology; OMIM, Online Mendelian Inheritance in Man.

* Corresponding author.
** Correspondence to: L. Chen, Key Laboratory of Systems Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China.
E-mail addresses: yangzhiyuansibs@gmail.com (Z. Yang), yanzhang01@sibs.ac.cn (Y. Zhang), lnchen@sibs.ac.cn (L. Chen).

http://dx.doi.org/10.1016/j.gene.2015.07.004
0378-1119/

produce a sense of skin pain in the cerebral cortex (Park et al., 2003). In addition, NMRs have the other characteristics, such as poor vision and hypoxia tolerance (Larson and Park, 2009; Peterson et al., 2012). These specific characteristics of NMRs make them useful models for investigating mechanisms of cancer, aging, and pain.

Cancer is a group of polygenic diseases caused by uncontrolled proliferation and differentiation (Ou et al., 2014). Cancer cells can spread to the surrounding normal tissues and result in large consumption of nutrients *in vivo* (Oberley et al., 2014). It is widely believed that the uncontrolled proliferation of cancer cell is activated by mutations of oncogenes or abnormal expressions of cancer-related genes (Kim et al., 2014; Monteferrario et al., 2014). During DNA transcription, some replacements or deletions may happen in the DNA sequence and result in mutation of translated protein. These translational errors can be accumulated slowly over time and turn a normal cell into a cancer cell eventually. Hence, indentifying the mutations of targeted proteins is an important task in cancer research. In fact, most of the human diseases do not depend on the proteins encoded by one gene, but depend on the protein–protein interaction (PPI) (Sun et al., 2014a). These interactions eventually form the large networks embedded in the organisms, that is, the metabolic networks and the signaling networks. These biological networks not only show the common properties of networks, but also have its unique characteristics, such as forming gene modules in the network (Chakraborty et al., 2014; Liu et al., 2014).

The recently updated version of NMR genome (Keane et al., 2014) provides great opportunities for investigating the special features of NMR and for improving its physiological and pathological studies. Several genes have been previously suggested to be related to some of these special features, *e.g.*, the p16 gene, which may be involved in the anti-cancer mechanisms of NMR (Maciak and Michalak, 2014). However, research at the network level is still a vacancy, which may provide important clues to reveal the physiological mechanisms for the special features of NMR.

In this study, a comparative genomics was carried out to explore the genes that are either common, or absent in NMR. We then analyzed known cancer-related genes that are absent or unique in NMR. In addition, we compared NMR with other 11 mammals to identify NMR-specific single amino acid changes (SAACs) among orthologous proteins of these organisms. Interestingly, a significant number of genes with these changes were found to be enriched in the pathways related to the exceptional characteristics of NMR, many of which have been previously reported to be associated with various cancers. Overall, our data not only help unveil the possible cancer resistance mechanisms of NMR but provide insights into identifying new cancer-related genes.

2. Methods

2.1. Genome, database and resources

In this study, we only focused on the annotated protein-coding genes. The protein sequence datasets of NMR and other 11 mammals: *Homo sapiens* (human), *Mus musculus* (mouse), *Rattus Norvegicus* (rat), *Callithrix jacchus* (marmoset), *Equus caballus* (horse), *Pongo abelii* (Sumatran orangutan), *Macaca mulatta* (macaque), *Pan troglodytes* (chimpanzee), *Canis familiaris* (dog), *Bos taurus* (bovine), *and Felis catus* (cat), were obtained from the UniProt database (http://www.uniprot.org/). For those genes with alternative splicing variants, sequences with the longest length were chosen to represent the gene-encoding protein sequences. For example, a total of 21,487 proteins corresponding to their genes were finally obtained for NMR (UniProt release 2013_11). The rat protein dataset was used as a representative to show the details of the comparison between NMR and other mammals. The files containing the whole pathways of rat/mouse/human were downloaded from the KEGG database (http://www.genome.jp/kegg/pathway.html).

2.2. Analysis of specific genes of NMR

To analyze the orthologous gene pairs between NMR and other mammals, we used the complete set of annotated proteins of one organism as queries to search for orthologs in the other species *via* BLASTP with a cut-off of E-value ≤ 1e-6. Orthologous genes were further defined as bidirectional best hits.

On the basis of the identified orthologs between NMR and other mammals, we singled out both NMR-specific genes and NMR-missing genes. If a gene is present in NMR but could not be found in other 11 mammals, it was considered as a NMR-specific gene. Similarly, if a gene is present in all 11 mammals but absent in NMR, it was assumed to be a candidate of NMR-missing gene. We then validated these candidate genes by the published RNA-seq datasets of NMR (Fang et al., 2014). To avoid mis-annotation, the *Cavia porcellus* (guinea pig), a well-annotated model organism that is extremely close to NMR in phylogenetic tree, was further used as a reference to check the proteins of these candidate genes (Deweerdt, 2014).

The ProtPraram tool (http://web.expasy.org/protparam/) was used to calculate the physicochemical properties of proteins, such as molecular weight and isoelectric point (pI). To investigate the possible mechanisms of the anti-cancer aspects of NMR, these genes were then searched against the Atlas of Genetics and Cytogenetics in Oncology and Haematology (AGCOH) database (Huret et al., 2013) (http://atlasgeneticsoncology.org/index.html) and Tumor Suppressor Gene (TSgene) database (Zhao et al., 2013) (http://bioinfo.mc.vanderbilt.edu/TSGene/) to identify cancer-related NMR-missing genes.

2.3. Analysis of single amino acid changes of highly conserved orthologous genes

In this study, the rat protein dataset was used as a representative to show the details of the comparison between NMR and other mammals. Although orthologous genes are considered as common genes between rat and NMR, SAACs have been previously reported for certain proteins in NMR. For example, the NMR HAS2 (hyaluronan synthase 2) protein has "N178S" and "N301S" peculiar mutations when compared to several other organisms. Such variations could result in the lack of activity of HA-degrading enzymes and high molecular mass of hyaluronan, responsible for the unusual cancer resistance in NMR (Tian et al., 2013). Thus, it would be interesting to investigate whether such events are also common for many other genes and their potential impacts on protein function.

In this study, we only focused on highly conserved orthologous genes in all examined organisms, with a cut-off of the percentage of identical matches ≥90% and the number of mismatches ≤5 based on BLASTP results. Multiple alignments of orthologous proteins from NMR and other 11 mammals were analyzed using ClustalW with default parameters. The NMR-specific SAAC event was assigned for a protein if all the other mammals have the same amino acid, which is different from the NMR one at certain position in the alignment. The NMR genes with SAACs were also searched against the NMR genome dataset *via* TBLASTN for the verification of such amino acid mutations.

2.4. KEGG pathway analysis of SAAC genes

We carried out pathway enrichment analysis for SAAC genes using the KEGG database resource, which contained 274 curated pathway maps on protein–protein interactions (up to November 2013). The enrichment p-value of these genes in each pathway was calculated by hypergeometric test. The Benjamini–Hochberg procedures (Benjamini and Hochberg, 1995) were conducted to calculate the False Discovery Rate (FDR) for multiple comparisons. Several cancer-related pathways with p-value ≤ 0.05 were selected for further analysis.