

# Genome sequencing reveals insights into physiology and longevity of the naked mole rat

Eun Bae Kim<sup>1\*</sup>, Xiaodong Fang<sup>2\*</sup>, Alexey A. Fushman<sup>1\*</sup>, Zhiyong Huang<sup>2\*</sup>, Alexei V. Lobanov<sup>3</sup>, Lijuan Han<sup>2</sup>, Stefano M. Marino<sup>3</sup>, Xiaoqing Sun<sup>2</sup>, Anton A. Turanov<sup>3</sup>, Pengcheng Yang<sup>2</sup>, Sun Hee Yim<sup>3</sup>, Xiang Zhao<sup>2</sup>, Marina V. Kasaikina<sup>3</sup>, Nina Stoletzki<sup>3</sup>, Chunfang Peng<sup>2</sup>, Paz Polak<sup>3</sup>, Zhiqiang Xiong<sup>2</sup>, Adam Kiezun<sup>3</sup>, Yabing Zhu<sup>2</sup>, Yuanxin Chen<sup>2</sup>, Gregory V. Kryukov<sup>3,4</sup>, Qiang Zhang<sup>2</sup>, Leonid Peshkin<sup>5</sup>, Lan Yang<sup>2</sup>, Roderick T. Bronson<sup>6</sup>, Rochelle Buffenstein<sup>7</sup>, Bo Wang<sup>2</sup>, Changlei Han<sup>2</sup>, Qiye Li<sup>2</sup>, Li Chen<sup>2</sup>, Wei Zhao<sup>2</sup>, Shamil R. Sunyaev<sup>3,4</sup>, Thomas J. Park<sup>8</sup>, Guojie Zhang<sup>2</sup>, Jun Wang<sup>2,9,10</sup> & Vadim N. Gladyshev<sup>1,3,4</sup>

**The naked mole rat (*Heterocephalus glaber*) is a strictly subterranean, extraordinarily long-lived eusocial mammal<sup>1</sup>. Although it is the size of a mouse, its maximum lifespan exceeds 30 years, making this animal the longest-living rodent. Naked mole rats show negligible senescence, no age-related increase in mortality, and high fecundity until death<sup>2</sup>. In addition to delayed ageing, they are resistant to both spontaneous cancer and experimentally induced tumorigenesis<sup>3,4</sup>. Naked mole rats pose a challenge to the theories that link ageing, cancer and redox homeostasis. Although characterized by significant oxidative stress<sup>5</sup>, the naked mole rat proteome does not show age-related susceptibility to oxidative damage or increased ubiquitination<sup>6</sup>. Naked mole rats naturally reside in large colonies with a single breeding female, the ‘queen’, who suppresses the sexual maturity of her subordinates<sup>7</sup>. They also live in full darkness, at low oxygen and high carbon dioxide concentrations<sup>8</sup>, and are unable to sustain thermogenesis<sup>9</sup> nor feel certain types of pain<sup>10,11</sup>. Here we report the sequencing and analysis of the naked mole rat genome, which reveals unique genome features and molecular adaptations consistent with cancer resistance, poikilothermy, hairlessness and insensitivity to low oxygen, and altered visual function, circadian rhythms and taste sensing. This information provides insights into the naked mole rat’s exceptional longevity and ability to live in hostile conditions, in the dark and at low oxygen. The extreme traits of the naked mole rat, together with the reported genome and transcriptome information, offer opportunities for understanding ageing and advancing other areas of biological and biomedical research.**

We applied a whole-genome shotgun strategy to sequence the genome of an individual male naked mole rat (NMR) (Table 1 and Supplementary Tables 1–3). The sequencing depth of 98.6% of the

genome assembly was more than 20-fold (Supplementary Figs 1–4). The mitochondrial genome was also assembled. Approximately 25% of the NMR genome was represented by transposon-derived repeats, which is lower than in other mammals (40% in human, 37% in mouse, and 35% in rat genomes) (Supplementary Tables 4 and 5, Supplementary Figs 5–7). The predicted NMR gene set included 22,561 genes (Table 1 and Supplementary Table 6), which is comparable to other mammals (22,389 in human, 23,317 in mouse, and 22,841 in rat). Of these, 21,394 (94.8%) genes were transcribed (based on the RNA-seq data for seven organs). More than 98% of NMR genes could be functionally annotated using homology approaches (Supplementary Table 7), and the quality of predicted genes was comparable to that of well-annotated mammalian genomes (Supplementary Tables 6 and 8 and Supplementary Fig. 8).

Most of the NMR genome (93%) showed synteny to human, mouse or rat genomes (Supplementary Table 9), and pairwise comparisons suggested a relatively low rate of NMR genome rearrangements after the split from the murid common ancestor. We defined common synteny blocks in human, mouse, rat and NMR genomes and identified segmental duplications and lineage-specific insertions and deletions (Supplementary Tables 10 and 11 and Supplementary Fig. 9). By analysing single-copy orthologous groups, we constructed a phylogenetic tree involving the NMR and other mammals (Fig. 1). As expected, the NMR placed within rodents and its ancestor split from the ancestor of rats and mice approximately 73 million years ago, whereas the ancestor of NMR, mouse and rat split from rabbit approximately 86 million years ago. Thus, in spite of some exceptional traits, the overall properties of the NMR genome appeared to be similar to those of other mammals.

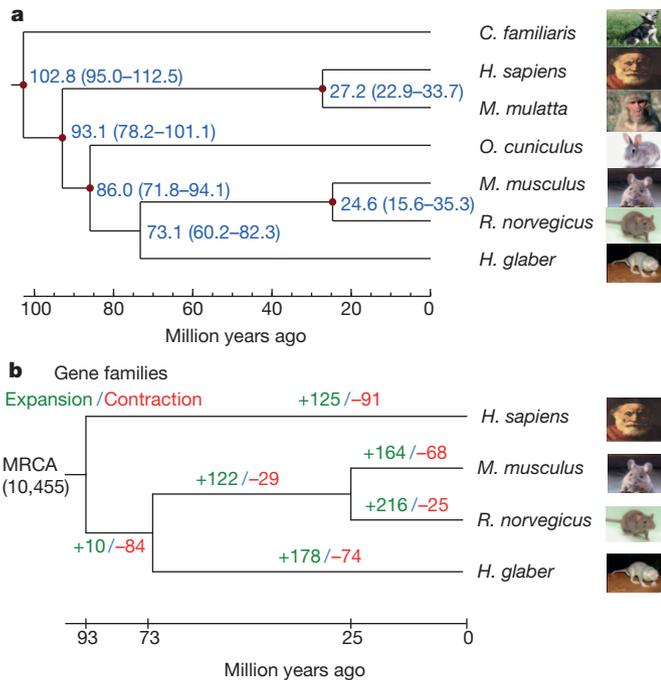
Lineage-specific gene family expansions may be associated with the emergence of specific functions and physiology. Compared to other mammals, the NMR showed a moderate number of gene families under expansion and contraction (Fig. 1b), including 96 NMR lineage-specific gene families (Fig. 2). Analysis of syntenic regions identified 750 gained and 320 lost NMR genes (Supplementary Tables 12–14). At least 75.5% of genes gained showed evidence of transcription, and the lost genes were enriched for ribosome and nucleoside biosynthesis functions (Supplementary Table 15). We also identified 244 pseudogenes, containing 183 frameshift and 119 premature termination events (Supplementary Tables 16 and 17). Functional categories enriched for pseudogenes included olfactory receptor activity (GO:0004984,  $P < 0.001$ , Fisher’s exact test, 36 genes), visual perception (GO:0007601,  $P = 0.015$ , *CRB1*, *CRYBB3*, *GNAT2*, *GRK7*, *GUCA1B* and *PDE6H*), spermatogenesis

**Table 1 | Global statistics of the NMR genome**

Sequencing	Insert size (bp)	Total data (Gb)	Sequence coverage (fold)
Paired end libraries	170–800	126.52	47
	2–20 × 10 <sup>3</sup>	120.66	45
	Total	247.18	92
Assembly	N50 (kb)	Longest (kb)	Size (Gb)
Contigs	19.3	179	2.45
Scaffolds	1,585	7,787	2.66
Annotation	Number	Total length (Mb)	Percentage of the genome
Repeats	3,090,116	666.7	25
Genes	22,561	722.3	27.1
CDS	181,641	32.5	1.2

<sup>1</sup>Department of Bioinspired Science, Ewha Womans University, Seoul, 120-750, Korea. <sup>2</sup>BGI-Shenzhen, Shenzhen, 518083, China. <sup>3</sup>Division of Genetics, Department of Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>4</sup>Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. <sup>5</sup>Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>6</sup>Rodent Histopathology Laboratory, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>7</sup>Department of Physiology and The Sam and Ann Barshop Institute for Longevity and Aging Studies, University of Texas Health Science Center, San Antonio, Texas 78245, USA. <sup>8</sup>Department of Biological Sciences, University of Illinois at Chicago, Chicago, Illinois 60607, USA. <sup>9</sup>Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, DK-2200 Copenhagen N, Denmark. <sup>10</sup>Department of Biology, University of Copenhagen, Copenhagen, DK-2200 Copenhagen N, Denmark.

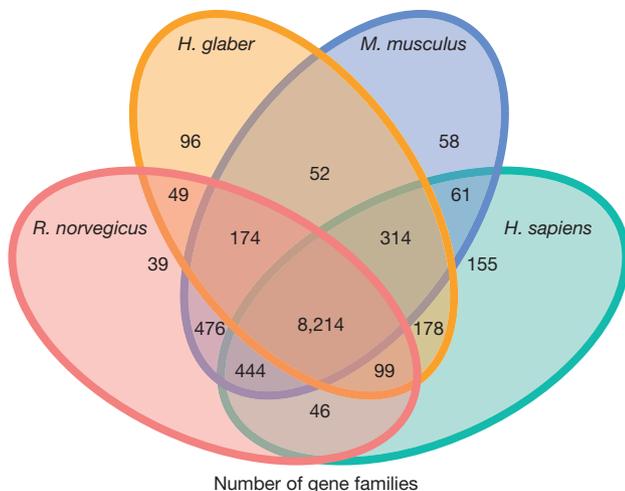
\*These authors contributed equally to this work.



**Figure 1 | Relationship of the NMR to other mammals.** **a**, Estimation of the time of divergence (with error range shown in parentheses) of the NMR and six other mammals based on orthology relationship. Distances are shown in millions of years. **b**, Expansion and contraction in gene families. Numbers designate the number of gene families that have expanded (green) and contracted (red) since the split from the common ancestor. The most recent common ancestor (MCRA) has 10,455 gene families.

(GO:0007283,  $P = 0.044$ , *ADAM29*, *ADC*, *CCIN*, *CCT6B*, *DEDD*, *OAZ3* and *SHBG*), and possibly RING domain (SM00184,  $P = 0.142$ , *CNOT4*, *KCNRG*, *RNF5*, *TRIM17*, *TRIML1* and *ZSWIM2*). The enrichment in the visual perception category appears to underlie the evolution of poor vision in the NMR, whereas many RING-domain-containing proteins act as ubiquitin ligases<sup>12</sup>. The levels of ubiquitinated proteins in NMRs are lower than in mice and, unlike those in mice, do not change significantly with age<sup>6</sup>.

Identification of genes that have undergone positive selection in the NMR lineage can provide useful pointers to the evolution of its unique



**Figure 2 | Common and unique NMR gene families.** This Venn diagram shows unique and overlapping gene families in the NMR (*H. glaber*), rat (*R. norvegicus*), mouse (*M. musculus*) and human (*H. sapiens*).

traits. 45 genes (0.4%) were identified as positively selected in the NMR lineage at the false discovery rate of 0.01 and 141 genes (1.2%) at the false discovery rate of 0.05 (Supplementary Table 18). 12 out of the 45 genes (corresponding to the false discovery rate of 0.01) passed a strict manual inspection for alignment quality. In comparison, 0.7% of genes were predicted to be positively selected in the human lineage from high-quality alignments and using Rom correction for multiple testing<sup>13</sup>. Interestingly, our set included *TEP1*, encoding a telomerase component, and *TERF1*, a telomeric repeat binding factor identified at the false discovery rate of 0.05 (Supplementary Fig. 10). The *TERF1* gene product is one of six proteins contributing to the shelterin complex, which shapes and protects telomeres<sup>14</sup> and has been proposed to regulate telomere length<sup>15</sup>.

To gain further insights into biological processes that underlie the exceptional traits of the NMR, we identified 39 NMR proteins containing 45 amino acid residues unique among orthologues present in 36 vertebrate genomes (Supplementary Table 19). This gene set included cyclin E1 (*CCNE1*), uncoupling protein 1 (*UCP1*) and  $\gamma$ -crystallin (*CRYGS*), which are associated with the G1/S transition during the cell cycle, thermogenesis and visual function, respectively. Other noteworthy genes were *APEX1*, a multifunctional DNA repair enzyme, *RFC1*, replication factor C, and *TOP2A*, a DNA topoisomerase that controls the topologic states of DNA during transcription. This set also contained eight genes designated as cancer-related<sup>16</sup>. Finally, *TOP2A*, along with *TEP1* and *TERF1* from the set of positively selected genes, are part of a five-protein complex of alternate lengthening of telomere pathway<sup>17</sup>. Overall, these analyses point to altered telomerase function in the NMR, which may be related to its evolution of extended lifespan and cancer resistance.

We also identified 1.87 million heterozygous single-nucleotide polymorphisms (SNPs). This results in an estimated nucleotide diversity (mean per nucleotide heterozygosity) of  $7 \times 10^{-4}$ , which is much lower than in mouse and rat populations and is comparable to the nucleotide diversity observed in humans. Transition nucleotide changes were observed twice as often as transversions, indicating that variant calls reproduce the expected properties of natural variation in other mammals. This low level of nucleotide diversity may reflect a low effective size of NMR population, but may also be due to a high level of inbreeding, a reduced mutation rate or high efficiency of the repair systems. The variation of diversity along the genome was consistent with inbreeding in the NMR population. In protein-coding regions of the genome, our analysis identified 10,951 non-synonymous and 8,616 synonymous SNPs. Their ratio is much higher than in other studied organisms, including human, which appears to signal relaxation of purifying selection in the NMR, potentially as a consequence of reduced effective population size. Finally, we analysed the context dependency of NMR SNPs (Supplementary Fig. 11). Relative rates of nucleotide changes and nucleotide context dependencies were similar to those observed in human polymorphism, with the exception of a relative reduction of SNPs due to CpG mutations. This was caused by a combination of the relatively low CpG density in the NMR genome and a higher fraction of CpG dinucleotides within CpG islands compared to the human genome. CpG density was only 0.19 of that expected on the basis of the GC content, which is lower than in human, dog and panda genomes, but is similar to the mouse genome. However, in comparison to mouse, a higher fraction of CpG dinucleotides was concentrated in CpG islands. CpG dinucleotides within CpG islands contribute less to genetic variation because of their lower methylation rate and possibly also due to selection.

Long lifespan is a key feature of the NMR. To study ageing and longevity, we obtained RNA-seq data for brain, liver and kidney of newborn, young adult (4-year-old) and old adult (20-year-old) NMRs (Supplementary Table 20). In contrast to other mammals, few genes showed differential expression between 4- and 20-year-old NMRs, especially in the brain (Supplementary Tables 21–23). A recent study identified 33 underexpressed and 21 overexpressed genes in the human brain during ageing<sup>18</sup>. Of these, 32 genes did not show

consistent expression changes with ageing in NMRs, including 30 genes that had stable expression and two genes that changed in the opposite direction compared to human brain (Supplementary Table 21). For example, *CYP46A1* and *SMAD3* were downregulated in the human brain, but showed elevated expression in the NMR brain. The product of the *CYP46A1* gene is a mediator of cholesterol homeostasis that influences the tendency of A $\beta$  to aggregate. The product of *SMAD3* is a modulator of TGF- $\beta$  signalling, playing a role in cancer development by slowing down the rate of cell proliferation. Elevated expression of *SMAD3* in the NMR during ageing may help optimize the rate of cell death, protecting NMRs from cancer.

A previous meta-analysis of age-related gene expression in mice, rats and humans revealed 56 consistently overexpressed and 17 under-expressed genes<sup>19</sup>. However, many of these genes did not show the same expression changes, suggesting that different regulatory mechanisms may underlie NMR longevity (Supplementary Tables 22 and 23). For example, genes related to degradation of macromolecules, such as *GSTA1*, *DERL1* and *GNS*, were not upregulated with age in NMRs. We also found that genes encoding mitochondrial proteins (*NDUFB11*, *ATP5G3* and *UQCRCQ*) were not downregulated, consistent with stable maintenance of mitochondrial function during ageing. It is also of interest that *TERT* (telomerase reverse transcriptase) showed stable expression regardless of age (Supplementary Fig. 12). This finding is consistent with the role of the telomerase complex, highlighted by positive selection on *TEP1* and *TERF1*. Overall, transcriptome and sequence data revealed different (compared to humans, mice and rats) patterns of NMR genes, which may underlie longevity mechanisms in this animal.

Non-shivering thermogenesis is a major heat production process in mammals that mainly depends on the action of *UCP1*, one of the 39 vertebrate genes that changed uniquely in the NMR (Supplementary Table 19). *UCP1* featured changes in amino acids Gln 146, Arg 263, Trp 264 and Thr 303, with the latter two residues being subject to positive selection ( $P < 0.05$ , likelihood ratio test for the branch-site model,  $n = 30$ ) and Arg 263 and Trp 264 located in the conserved nucleotide binding motif (Fig. 3a). With Arg–Trp instead of the rigid Gly–Pro in the key regulatory site, *UCP1* is expected to lose the tight regulation by purine nucleotides as inhibitors and fatty acids as activators (Fig. 3b and c). The same loop also features two positively

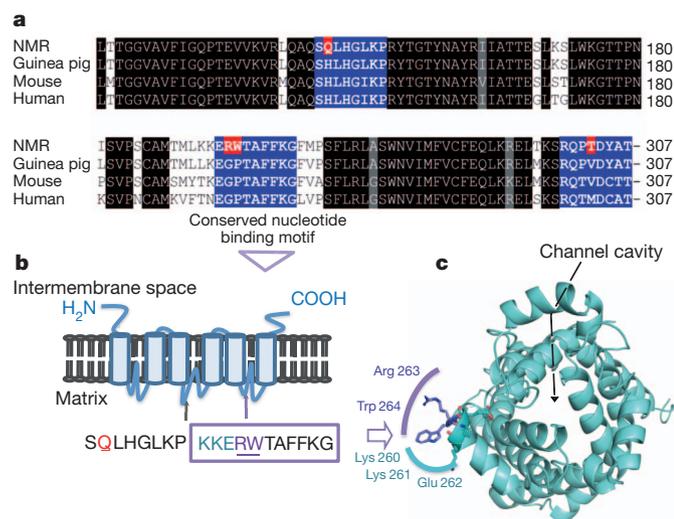
charged Lys residues followed by a negatively charged residue (also a unique combination), that should markedly affect the local electrostatic potential of *UCP1*. In addition, Gln 146 replaced a conserved His involved in proton transport, and the same mutation was shown to decrease proton conductance of *UCP1* fivefold<sup>20</sup>. Thr 303 is located in the carboxy-terminal motif (RqTxDCxT) required for binding purine nucleotides<sup>21</sup>. Taken together, these observations indicate a tight association of *UCP1* function with the unique thermoregulation of the NMR<sup>22</sup>.

In mammals, switches between light and dark periods affect synthesis of the hormone melatonin, which modulates sleep and circadian rhythms. NMRs live in a naturally dark habitat and their pineal glands, where melatonin is synthesized, are atrophied<sup>23</sup>, but we found that the genes involved in melatonin synthesis (*TPH1*, *TPH2*, *DDC*, *AANAT* and *ASMT*) are intact. Interestingly, the expression of genes involved in the final two steps of melatonin synthesis was very low (*AANAT*) or undetectable (*ASMT*) in the NMR brain regardless of age (Supplementary Table 24 and Supplementary Fig. 13). Moreover, two major mammalian melatonin receptors (*MTNRIA* and *MTNR1B*, encoding MT<sub>1</sub> and MT<sub>2</sub>, respectively) were inactivated by mutations that introduce premature stop signals (Supplementary Fig. 14). Synteny analyses showed that these pseudogenes corresponded to mouse *MTNRIA* and *MTNR1B*. Although melatonin signalling appears to be disrupted in the NMR, its circadian rhythms were maintained in terms of locomotor activity and body temperature when exposed to periodic light/dark changes<sup>24</sup>. Our finding is consistent with a previous report that MT<sub>1</sub>/MT<sub>2</sub> knockout mice maintained essentially normal circadian rhythms<sup>25</sup>. These mice also showed decreased insulin secretion<sup>25</sup>. Likewise, our transcriptome analysis of the NMR revealed decreased expression of genes involved in insulin/IGF-1 signalling in the liver compared to mice (Supplementary Fig. 15).

To explain the extraordinary resistance of the NMR to cancer<sup>3</sup>, a two-tier protective mechanism involving contact inhibition mediated by p16<sup>Ink4a</sup> and p27<sup>Kip1</sup> was proposed<sup>4</sup>. The involvement of p16<sup>Ink4a</sup> is unusual, since humans and mice show only contact inhibition mediated by p27<sup>Kip1</sup>. We analysed the gene locus and the transcriptome reads corresponding to tumour suppressors p16<sup>Ink4a</sup> and p19<sup>Arf</sup>. As in mice, the p16<sup>Ink4a</sup> transcript consists of three exons (Supplementary Fig. 16). However, sequence similarity in the last exon is low, and two early stop codons in the second exon were predicted to result in a shorter, 14-kDa protein (Supplementary Fig. 17). The four ankyrin repeats were, however, intact and Thr69, a residue important for CDK6 binding, was conserved, so the function of the protein may be partially preserved (Supplementary Fig. 18). The p19<sup>Arf</sup> transcript consists of two exons, but four stop codons in the second exon should lead to a shorter, 10-kDa protein (Supplementary Figs 19–21).

The NMR is also unique in that its skin and cutaneous C-fibres lack the neuropeptide Substance P, making this animal insensitive to certain types of pain<sup>10,11</sup>. Our analysis revealed the presence of intact *TAC1* encoding Substance P. However, the NMR had a deletion in the core promoter region highly conserved among mammals (Supplementary Fig. 22). Thus, this neurotransmitter appears to be functional but may be under unique regulation.

We further examined the molecular basis for poor visual function and small eyes in the NMR. Of the four vertebrate opsin genes (*RHO*, *OPN1LW*, *OPN1MW* and *OPN1SW*), two (*OPN1LW* and *OPN1MW*) were missing (Table 2); this distinguishes the NMR from other rodents with dichromatic colour vision, such as mice, rats and guinea pigs. However, the NMR has intact *RHO* (rhodopsin) and *OPN4* (melanopsin), supporting the presence of rod-dominated retinæ and the capacity to distinguish light/dark cues. Of about 200 genes associated with visual perception (GO:0007601) in humans and mice, almost 10% were inactivated or missing in the NMR (Table 2 and Supplementary Fig. 23). These mammalian genes participate in crystallin formation, phototransduction in the retina, retinal development, dark adaptation, night blindness and colour vision. For at least ten of these



**Figure 3 | Unique changes in UCP1 sequences and their roles in thermoregulation.** **a**, Alignment of mammalian UCP1 sequences. Amino acids unique to the NMR are highlighted in red, and conserved motifs in blue. **b**, Topology of UCP1. Regions affected in the NMR are highlighted. **c**, Structural model of UCP1. Location of the channel and the nucleotide-binding loop with altered sequences in the NMR are shown.

**Table 2 | Visual perception genes that are inactivated or are missing in the NMR genome**

Gene	Inactivation event	Time of gene loss	$\omega_0$ (average)	$\omega_1$ (other)	$\omega_2$ (NMR)	P-value
<i>RBP3</i>	F	NMR	0.121	0.085	0.232	1.187E-11
<i>ARR3</i>	F/S	NMR	0.420	0.260	0.912	6.263E-06
<i>PDE6C</i>	F	NMR	0.171	0.139	0.316	0.0001
<i>GUCA1B</i>	F	NMR	0.083	0.056	0.217	0.001
<i>GJA10</i>	F/S	NMR	0.308	0.248	0.524	0.002
<i>GUCY2E</i>	F	NMR	0.124	0.105	0.182	0.002
<i>CRYBA4</i>	S	NMR	0.055	0.036	0.123	0.001
<i>GNAT2</i>	F	NMR	0.055	0.039	0.108	0.017
<i>SLC24A1</i>	F	NMR	0.389	0.355	0.517	0.035
<i>CRYBB3</i>	S	NMR	0.071	0.048	0.122	0.037
<i>RP1L1</i>	S	NMR	0.448	0.424	0.513	0.186
<i>GRK7</i>	F	NMR	0.154	0.135	0.201	0.335
<i>PDE6H</i>	F	NMR	0.091	0.082	0.127	0.648
<i>EYS</i>	F/S	Ancestor	-	-	-	-
<i>GUCA1C</i>	F	Ancestor	-	-	-	-
<i>OPN1LW</i>	L	NMR	-	-	-	-
<i>OPN1MW</i>	L	NMR	-	-	-	-
<i>PRCD</i>	L	NMR	-	-	-	-
<i>RD3</i>	L	NMR	-	-	-	-

F, frameshift; S, premature stop codon; L, complete loss detected with synteny information. Ancestor/NMR indicates that the gene was lost in a rodent ancestor or the NMR, respectively. The rate ratio  $\omega$  of non-synonymous to synonymous substitutions was calculated by using human, mouse, rat and NMR sequences.  $\omega_0$  is the average ratio in all branches,  $\omega_1$  is the average ratio in non-NMR branches, and  $\omega_2$  is the ratio in the NMR branch. A small P-value indicates that the two-ratio model fits the data better than the one-ratio model.

genes, we observed relaxation of the functional constrain on NMR sequences by estimating the ratio of non-synonymous to synonymous substitutions, which corroborated the dysfunction of these genes. Inactivation of *CRYBA4*, a microphthalmia-related gene, may be associated with the small-sized eyes, whereas inactivation of *CRYBA4* and *CRYBB3* and a NMR-specific mutation in *CRYGS* (Supplementary Table 19) may be associated with abnormal eye morphology<sup>26</sup>. Thus, while some genes responsible for vision are preserved in the NMR, its poor visual function may be explained by deterioration of genes coding for various critical components of the visual system.

Further analysis revealed substantial divergence of the NMR nuclear receptor corepressor Hairless from other mammalian orthologues and the presence of amino acid replacements associated with the hairless phenotype, which is consistent with the lack of fur in NMRs (Supplementary Fig. 24). In addition, we found substantial sequence variation in the sweet taste receptor and lack of many bitter taste receptors common to other mammals (Supplementary Fig. 25 and 26). In particular, the NMR appears to lack the phenylthiocarbamide taste, a dominant genetic trait in humans, as well as several other common bitter tastes.

Air in NMR burrows is low in O<sub>2</sub> (~8%) and high in CO<sub>2</sub> (>10%) owing to many animals sharing a limited air supply and poor gas exchange through soil<sup>27</sup>. To cope with the low O<sub>2</sub> conditions, the NMR has developed adaptive circulatory (altered haemoglobin oxygen affinity) and metabolic functions, reducing metabolic rate and slowing down development<sup>1,8,28,29</sup>. To obtain insights into this adaptation, we examined gene expression changes in several tissues of NMR subjected to 8% O<sub>2</sub> for one week (Supplementary Tables 25-31 and Supplementary Fig. 27-30). Many changes associated with energy metabolism and redox control were observed. Sequence analysis of NMR hypoxia-induced factor 1 $\alpha$  (HIF1 $\alpha$ ) revealed a T407I exchange unique among mammals and located in the VHL-binding domain. Under normal oxygen conditions, VHL mediates ubiquitin-dependent degradation of HIF1 $\alpha$ . In addition, NMR VHL harbours V166I exchange at a functionally important site. These amino acid changes are consistent with relaxation of ubiquitin-dependent degradation of HIF1 $\alpha$ , and, thus, with adaptation to low oxygen conditions.

To summarize, sequencing and analysis of the NMR genome revealed numerous insights into the biology of this remarkable animal. In addition, this genome and the associated data sets offer the research

communities working in ageing, cancer, eusociality and many other areas a rich resource that can be mined in numerous ways to uncover the molecular bases for the extraordinary traits of this most unusual mammal. In turn, this information provides unprecedented opportunities for addressing some of the most challenging questions in biology and medicine, such as mechanisms of ageing, the role of genetic makeup in regulating lifespan, adaptations to extreme environments, hypoxia tolerance, thermogenesis, resistance to cancer, circadian rhythms, sexual development and hormonal regulation.

## METHODS SUMMARY

The NMR genome was sequenced on the Illumina HiSeq 2000 platform. The sequenced individual male NMR was from a captive breeding colony located at the University of Illinois, Chicago. The genome was assembled using SOAPdenovo. We obtained 2.5 Gb (gigabase pairs) contig sequences with N50 19.3 kb (kilobase pairs) and N90 4.7 kb, and 2.7 Gb scaffold sequences with N50 1.6 Mb (megabase pairs) and N90 0.3 Mb. (The N50 (or N90) contig size is the length of the smallest contig S in the sorted list of all contigs where the cumulative length from the largest contig to contig S is at least 50% (or 90%) of the total assembly length.) RNA-seq data (ageing and low O<sub>2</sub> experiments) were for animals from the same colony. See Supplementary Information for data analysis and additional details.

Received 14 June; accepted 5 September 2011.

Published online 12 October 2011.

- Edrey, Y. H., Park, T. J., Kang, H., Biney, A. & Buffenstein, R. Endocrine function and neurobiology of the longest-living rodent, the naked mole-rat. *Exp. Gerontol.* **46**, 116–123 (2011).
- Buffenstein, R. Negligible senescence in the longest living rodent, the naked mole-rat: insights from a successfully aging species. *J. Comp. Physiol. B* **178**, 439–445 (2008).
- Liang, S., Mele, J., Wu, Y., Buffenstein, R. & Hornsby, P. J. Resistance to experimental tumorigenesis in cells of a long-lived mammal, the naked mole-rat (*Heterocephalus glaber*). *Aging Cell* **9**, 626–635 (2010).
- Seluanov, A. et al. Hypersensitivity to contact inhibition provides a clue to cancer resistance of naked mole-rat. *Proc. Natl Acad. Sci. USA* **106**, 19352–19357 (2009).
- Andziak, B. et al. High oxidative damage levels in the longest-living rodent, the naked mole-rat. *Aging Cell* **5**, 463–471 (2006).
- Pérez, V. I. et al. Protein stability and resistance to oxidative stress are determinants of longevity in the longest-living rodent, the naked mole-rat. *Proc. Natl Acad. Sci. USA* **106**, 3059–3064 (2009).
- Jarvis, J. U. M. Eusociality in a mammal: cooperative breeding in naked mole-rat colonies. *Science* **212**, 571–573 (1981).
- Larson, J. & Park, T. J. Extreme hypoxia tolerance of naked mole-rat brain. *Neuroreport* **20**, 1634–1637 (2009).
- Buffenstein, R., Woodley, R., Thomadakis, C., Daly, T. J. & Gray, D. A. Cold-induced changes in thyroid function in a poikilothermic mammal, the naked mole-rat. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **280**, R149–R155 (2001).
- Park, T. J. et al. Selective inflammatory pain insensitivity in the African naked mole-rat (*Heterocephalus glaber*). *PLoS Biol.* **6**, e13 (2008).
- Smith, E. S., Blass, G. R., Lewin, G. R. & Park, T. J. Absence of histamine-induced itch in the African naked mole-rat and “rescue” by Substance P. *Mol. Pain* **6**, 29 (2010).
- Deshaies, R. J. & Joazeiro, C. A. P. RING domain E3 ubiquitin ligases. *Annu. Rev. Biochem.* **78**, 399–434 (2009).
- Schneider, A. et al. Estimates of positive Darwinian selection are inflated by errors in sequencing, annotation, and alignment. *Genome Biol. Evol.* **1**, 114–118 (2009).
- de Lange, T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.* **19**, 2100–2110 (2005).
- van Steensel, B. & Lange, T. Control of telomere length by the human telomeric protein TRF1. *Nature* **385**, 740–743 (1997).
- Higgins, M. E. et al. CancerGenes: a gene selection resource for cancer genome projects. *Nucleic Acids Res.* **35**, D721–D726 (2007).
- Bhattacharyya, S. et al. Telomerase-associated protein 1, HSP90, and topoisomerase IIA associate directly with the BLM helicase in immortalized cells using ALT and modulate its helicase activity using telomeric DNA substrates. *J. Biol. Chem.* **284**, 14966–14977 (2009).
- Hong, M. G., Myers, A. J., Magnusson, P. K. E. & Prince, J. A. Transcriptome-wide assessment of human brain and lymphocyte senescence. *PLoS ONE* **3**, e3024 (2008).
- de Magalhães, J. P., Curado, J. & Church, G. M. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics* **25**, 875–881 (2009).
- Bienengraeber, M., Echtay, K. S. & Klingenberg, M. H<sup>+</sup> transport by uncoupling protein (UCP-1) is dependent on a histidine pair, absent in UCP-2 and UCP-3. *Biochemistry* **37**, 3–8 (1998).
- Porter, R. K. Uncoupling protein 1: a short-circuit in the chemiosmotic process. *J. Bioenerg. Biomembr.* **40**, 457–461 (2008).

22. Woodley, R. & Buffenstein, R. Thermogenic changes with chronic cold exposure in the naked mole-rat (*Heterocephalus glaber*). *Comp. Biochem. Physiol. A* **133**, 827–834 (2002).
23. Quay, W. B. Pineal atrophy and other neuroendocrine and circumventricular features of the naked mole-rat, *Heterocephalus glaber* (Rüppell), a fossorial, equatorial rodent. *J. Neural Transm.* **52**, 107–115 (1981).
24. Riccio, A. P. & Goldman, B. D. Circadian rhythms of locomotor activity in naked mole-rats (*Heterocephalus glaber*). *Physiol. Behav.* **71**, 1–13 (2000).
25. Mühlbauer, E., Gross, E., Labucay, K., Wolgast, S. & Peschke, E. Loss of melatonin signalling and its impact on circadian rhythms in mouse organs regulating blood glucose. *Eur. J. Pharmacol.* **606**, 61–71 (2009).
26. Nikitina, N. V. *et al.* Postnatal development of the eye in the naked mole rat (*Heterocephalus glaber*). *Anat. Rec. A* **277A**, 317–337 (2004).
27. Bennett, N. C. & Faulkes, C. G. *African Mole-Rats: Ecology and Eusociality* (Cambridge University Press, 2000).
28. McNab, B. K. The influence of body size on the energetics and distribution of fossorial and burrowing mammals. *Ecology* **60**, 1010–1021 (1979).
29. Johansen, K., Lykkeboe, G., Weber, R. E. & Maloiy, G. M. Blood respiratory properties in the naked mole rat *Heterocephalus glaber*, a mammal of low body temperature. *Respir. Physiol.* **28**, 303–314 (1976).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank P. C. LaVinka for help in preparing NMR tissues and N. Estologa for help in preparing the manuscript. We acknowledge financial support from the WCU Program (R31-2008-000-10010-0), the NIH (AG038004, AG021518 and CA080946), the Shenzhen Municipal Government (ZYC200903240077A), the

National Natural Science Foundation of China (30725008) and the National Science Foundation (0744979).

**Author Contributions** V.N.G. conceived the study. T.J.P. carried out animal work. A.A.T, M.V.K. and S.H.Y. prepared samples. X.F., Z.H., L.H., X.S., P.Y., X.Z., C.P., Z.X., Y.Z., Y.C., Q.Z., L.Y., B.W., C.H., Q.L., L.C., W.Z., G.Z. and J.W. performed genome sequencing and assembly. X.F., G.Z. and J.W. supervised genome sequencing and assembly. E.B.K., X.F., A.A.F., Z.H., A.V.L., S.M.M., L.P., G.Z. and V.N.G. performed genome and transcriptome analyses. Z.H., N.S., P.P., A.K. and S.R.S. carried out genetic analyses. G.V.K., R.T.B. and R.B. discussed the data. All authors contributed to data interpretation. V.N.G. wrote the paper with significant contributions from E.B.K., X.F., A.A.F., Z.H., S.R.S. and G.Z., and input from all authors.

**Author Information** The NMR whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number AFSB00000000. The version described in this paper is the first version, AFSB01000000. The mitochondrial sequence has been deposited at GenBank under the accession number JN242813. All short-read data have been deposited into the Short Read Archive (<http://www.ncbi.nlm.nih.gov/sra>) under the accession number SRA030468. Raw sequencing data of the transcriptome have been deposited in the Gene Expression Omnibus with the accession number GSE30337. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at [www.nature.com/nature](http://www.nature.com/nature). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to G.Z. ([zhanggj@genomics.org.cn](mailto:zhanggj@genomics.org.cn)), J.W. ([wangj@genomics.org.cn](mailto:wangj@genomics.org.cn)) or V.N.G. ([vgladyshev@rics.bwh.harvard.edu](mailto:vgladyshev@rics.bwh.harvard.edu)).