



**Molecular Evolution of the SARS Coronavirus
During the Course of the SARS Epidemic in China**
The Chinese SARS Molecular Epidemiology
Consortium, *et al.*
Science **303**, 1666 (2004);
DOI: 10.1126/science.1092002

***The following resources related to this article are available online at
www.sciencemag.org (this information is current as of April 9, 2007):***

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/303/5664/1666>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/1092002/DC1>

This article **cites 25 articles**, 15 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/303/5664/1666#otherarticles>

This article has been **cited by** 29 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/cgi/content/full/303/5664/1666#otherarticles>

This article appears in the following **subject collections**:

Virology

<http://www.sciencemag.org/cgi/collection/virology>

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

Molecular Evolution of the SARS Coronavirus During the Course of the SARS Epidemic in China

The Chinese SARS Molecular Epidemiology Consortium*

Sixty-one SARS coronavirus genomic sequences derived from the early, middle, and late phases of the severe acute respiratory syndrome (SARS) epidemic were analyzed together with two viral sequences from palm civets. Genotypes characteristic of each phase were discovered, and the earliest genotypes were similar to the animal SARS-like coronaviruses. Major deletions were observed in the Orf8 region of the genome, both at the start and the end of the epidemic. The neutral mutation rate of the viral genome was constant but the amino acid substitution rate of the coding sequences slowed during the course of the epidemic. The spike protein showed the strongest initial responses to positive selection pressures, followed by subsequent purifying selection and eventual stabilization.

Severe acute respiratory syndrome (SARS) first emerged in Guangdong Province, China. Subsequently, the SARS coronavirus (SARS-CoV) was identified as the causative agent (1–5). It remains a challenge to establish the

relationship between observed genomic variations and the biology of SARS (4–8). Recent molecular epidemiological studies have identified characteristic variant sequences in SARS-CoV for tracking disease transmission (7, 9–11). Evidence suggests that SARS-CoV emerged from nonhuman sources (8, 12). In this study, we sought epidemiological and genetic evidence for viral adaptation to human beings through molecular investigations of the characteristic viral lineages found in China (13).

On the basis of epidemiological investigations (14), we divided the course of the epidemic into early, middle, and late phases (Fig. 1). The early phase is defined as the period from the first emergence of SARS to the first documented superspreader event (SSE) (13). The middle phase refers to the ensuing events up to the first cluster of SARS cases in a hotel (Hotel M) in Hong Kong (15). Cases following this cluster fall into the late phase.

The early phase was initially characterized by a series of seemingly independent cases. Eleven index cases that had arisen locally in the absence of any contact history were identified from different geographical locations within Guangdong Province (fig. S1). This phenomenon was observed from

the retrospectively identified SARS index patient from the city of Foshan (onset date, 16 November 2002) (13) through to an index patient from the city of Dongguan (onset date, 10 March 2003). All of these cases were confined to regions directly west of Guangzhou, the capital city of Guangdong Province, and to the city of Shenzhen in the south, with no cases being reported to the north or east of Guangzhou (Fig. 1) (fig. S1). This region, the Pearl River Delta, has enjoyed rapid economic development since the late 1970s, leading to the adoption of culinary habits requiring exotic animals. Seven of these 11 cases had documented contact with wild animals. In contrast to the apparently independent seeding of the earliest cases, the rest of the epidemic was characterized by SSEs and clusters of cases that were epidemiologically linked (Fig. 1) (fig. S1) (10, 11, 13, 15, 16).

The first major SARS outbreak occurred in a hospital, HZS-2, in the city of Guangzhou, beginning on 31 January 2003 where an SSE was identified to be associated with more than 130 primary and secondary infections, of which 106 were hospital-acquired cases. Doctor A, a nephrologist who worked in this hospital, visited Hong Kong and stayed in Hotel M on 21 February 2003. Other visitors to the hotel later became infected with SARS-CoV (13, 15). This led to the transmission of SARS to Vietnam, Canada, Singapore, and the United States (17) with two further SSEs in Hong Kong, each resulting in the virus being transmitted to >100 contacts (10, 16).

Genomic sequence data for SARS-CoV were largely derived from isolates linked to the Hotel M cluster (6), hence they were predominantly from the late phase of the epidemic. We determined 29 SARS-CoV genomic sequences obtained from 22 patients from Guangdong Province with disease onset dates in all three phases of the epidemic, and from two patients from the late phase in Hong Kong. To eliminate mutational noise, we assumed that sequence variants associated with common ancestry, but not arising in cell culture, should be seen in multiple isolates (7). Meanwhile, critical genomic variations or complete genome sequences of certain virus isolates were verified by sequencing the reverse transcription polymerase chain reaction (RT-PCR) products derived directly from patient specimens (14). The genomic sequences obtained were compared with 32 human SARS-CoV sequences and two SARS-like coronavirus sequences from Himalayan palm civets (*Paguma larvata*) available at GenBank as of the end of September 2003 (Fig. 2).

Only two major genotypes predominated during the early phase of the epidemic. Five isolates were found to contain a 29-nucleotide (nt) sequence that is absent in most of the publicly available SARS-CoV

*Epidemiology group: Jian-Feng He,^{1†} Guo-Wen Peng,¹ Jun Min,² De-Wen Yu,¹ Wen-Jia Liang,¹ Shu-Yi Zhang,³ Rui-Heng Xu,^{1‡} Virology group: Huan-Ying Zheng,^{1†} Xin-Wei Wu,^{4†} Jun Xu,⁵ Zhan-Hui Wang,⁶ Ling Fang,¹ Xin Zhang,¹ Hui Li,¹ Xin-Ge Yan,¹ Jia-Hai Lu,¹ Zhi-Hong Hu,⁷ Ji-Cheng Huang,¹ Zhuo-Yue Wan,^{1‡} Jin-Lin Hou,^{6‡} Jin-Yan Lin,^{1‡} Molecular biology group: Huai-Dong Song,^{8†} Sheng-Yue Wang,^{9†} Xiang-Jun Zhou,^{10†} Guo-Wei Zhang,⁸ Bo-Wei Gu,⁸ Hua-Jun Zheng,⁹ Xiang-Lin Zhang,⁹ Mei He,¹⁰ Kui Zheng,¹ Bo-Fei Wang,⁹ Gang Fu,⁹ Xiao-Ning Wang,⁶ Sai-Juan Chen,⁸ Zhu Chen,^{8,9‡} Data analysis group: Pei Hao,^{11,12†} Hua Tang,^{13†} Shuang-Xi Ren,^{9†} Yang Zhong,^{12,14†} Zong-Ming Guo,¹² Qi Liu,¹¹ You-Gang Miao,¹¹ Xiang-Yin Kong,¹¹ Wei-Zhong He,¹² Yi-Xue Li,^{11,12} Chung-I Wu,^{13‡} Guo-Ping Zhao,^{9,11‡} Chinese University of Hong Kong group: Rossa W. K. Chiu,^{15†} Stephen S. C. Chim,¹⁵ Yu-kwan Tong,¹⁵ Paul K. S. Chan,¹⁶ John S. Tam,¹⁶ Y. M. Dennis Lo,^{15‡}

¹Guangdong Center for Disease Control and Prevention, 176 Xingangxi Road, Guangzhou 510300, Guangdong, China. ²Second Affiliated Hospital of Sun Yat-Sen University, 107 Yanjiang West Road, Guangzhou 510120, Guangdong, China. ³Institute of Zoology, Chinese Academy of Sciences, 19 Zhongguancun Road, Haidian, Beijing 100080, China. ⁴Guangzhou Center for Disease Control and Prevention, 23 Third Zhongshan Road, Guangzhou 510080, Guangdong, China. ⁵Guangzhou Institute of Respiratory Disease, 151 Yanjiang West Road, Guangzhou 510120, Guangdong, China. ⁶Department of Infectious Diseases and Molecular Immunology, Nanfang Hospital, First Medical University of PLA, Guangzhou 510515, Guangdong, China. ⁷Wuhan Institute of Virology, Chinese Academy of Sciences, 44 Xiaohongshan, Wuhan 430071, Hubei, China. ⁸State Key Laboratory for Medical Genomics/Pôle Sino-Français de Recherche en Sciences du Vivant et Génomique, Ruijin Hospital Affiliated to Shanghai Second Medical University, 197 Rui Jin Road II, Shanghai 200025, China. ⁹Chinese National Human Genome Center at Shanghai, 250 Bi Bo Road, Zhang Jiang High Tech Park, Shanghai 201203, China. ¹⁰School of Pharmacy, Shanghai Jiaotong University, 1954 Huashan Road, Shanghai 200030, China. ¹¹Bioinformatics Center/Institute of Plant Physiology and Ecology/Health Science Center,

Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yue Yang Road, Shanghai 200031, China. ¹²Shanghai Center for Bioinformatics Technology, 100 Qinzhou Road, Shanghai 200235, China. ¹³Department of Ecology and Evolution, University of Chicago, 1101 E. 57th Street, Chicago, IL 60637, USA. ¹⁴School of Life Sciences, Fudan University, Shanghai 200433, China. ¹⁵Department of Chemical Pathology, ¹⁶Department of Microbiology, Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, New Territories, Hong Kong Special Administrative Region, China.

†These authors contributed equally to this work.

‡Corresponding authors for each group.

§Corresponding author for overall work. E-mail: gzhao@sibs.ac.cn

sequences, whereas another four isolates showed a previously unreported 82-nt deletion in the same region of the genome, Orf8 (18) (fig. S2 and table S1). The former sequence is represented by the GZ02 isolate [all GenBank accession numbers are listed in (14)] and is used as the reference for annotation throughout this study. All of the isolates exhibiting this sequence (GZ02, HGZ8L1-A, HSZ-A, HSZ-B, and HSZ-C; Fig. 2) were obtained from patients with contact histories traceable to some of the earliest independent cases in Guangzhou and were not detected in any of the later isolates. It is noteworthy that this sequence with the 29-nt segment is identical to the genomic sequence of coronaviruses isolated from animals in a Shenzhen live animal market (8).

Three of the SARS-CoV genome sequences (ZS-A, ZS-B, and ZS-C; Fig. 2) with the 82-nt deletion were obtained from samples of very early cases from Zhongshan city. This 82-nt deletion was further confirmed by RT-PCR directly on an additional stool sample. A sequence with an identical 82-nt deletion has also been observed in coronaviruses isolated from farmed civets in Hubei Province, China (19). It is thus interesting to note that both sequences of the early phase were identified from other mammalian hosts. They provided a link to support the notion that early human infection of SARS-CoV may have originated from wild animals (8, 12).

In contrast to the early phase, a SARS-CoV sequence with the 29-nt deletion was observed during the middle phase that dominated the viral population for the rest of the epidemic (4, 5, 7). Although this shift in genome size might be due to chance, deletion

events appeared to be overrepresented in the Orf8 region. A fourth sequence with the 82-nt deletion was obtained from a Guangzhou patient (HGZ8L1-B), who was infected in the same ward as one of the patients where the longest sequence was obtained (HGZ8L1-A) (see above). Furthermore, a lung biopsy of a patient from the middle phase was found to contain two SARS-CoV genotypes, with the 29-nt and the 82-nt deletions, respectively (fig. S3). Remarkably, another genotype with a 415-nt deletion resulting in the loss of the whole Orf8 region was isolated and confirmed in two Hong Kong patients with disease onset from mid-May 2003 (Fig. 2) (fig. S2) (20).

Because the majority of deletions observed in the SARS-CoV genome occurred in the Orf8 region with no apparent effect on the survival of the virus, it is tempting to suggest that this region is either noncoding or coding for a functionally unimportant putative protein (table S1). On the other hand, it is interesting to note that antiparallel reverse symmetrical sequences were readily predicted around the deletion sites (fig. S2), which might account for the high deletion rates in this region. Whether such hairpin structures actually play a role in regulating either RNA replication or mRNA transcription in SARS-CoV is a subject for future studies.

Besides the deletion variants, 299 single-nucleotide variations (SNVs) were detected among the 63 sequences. Eighty-five of these variant loci were seen in more than one of the human SARS-CoV sequences. Among them, 52 were predicted to cause amino acid changes (nonsynonymous variations) (table S2). When the epidemiologically determined transmission paths and SNV genotype data

are combined, markers for genotypes characteristic of different lineages are evident (Fig. 2) (table S2).

Viruses of the early phase have the characteristic motif of G:A:C:G:C at the GZ02 reference nucleotide residues 17,564, 21,721, 22,222, 23,823, and 27,827, with the bold SNVs matching the C:G:C:C motif identified previously (7) (Fig. 2). This motif is shared by almost all early Guangzhou and Zhongshan isolates together with the animal SARS-like coronavirus isolates (SZ3 and SZ16) (8). Along with the disappearance of viruses containing the 29-nt segment, the middle phase of the epidemic was characterized by the occurrence of genotypes with the G:A:C:T:C motif (Fig. 2). All of the middle-phase genotypes demonstrate this common motif but can be further classified into two variant groups on the basis of other SNVs (table S2). One group was represented by the isolates related to the Hospital HZS-2 outbreak (HZS2-A, HZS2-B, HZS2-C, HZS2-D, HZS2-E, and HGZ8L-2). The other group was represented by the Hong Kong CUHK-W1 isolate that originated from Shenzhen (9) together with the early Beijing isolates BJ01, BJ02, and BJ03, traceable to Guangdong. The transition between the characteristic motifs of the early and middle phases represented a G→T transversion at nucleotide residue 23,823 and is predicted to cause an Asp → Tyr change at amino acid residue 778 of the spike (S) protein (fig. S4).

An additional A→G transition at nucleotide 21,721 (Fig. 2) (fig. S4) was identified in one isolate from a secondarily infected patient from Hospital HZS-2 with disease onset on 7 February 2003 (HZS2-Fc) (Fig. 2). This sequence was additionally confirmed by direct sequencing of the RT-PCR product from

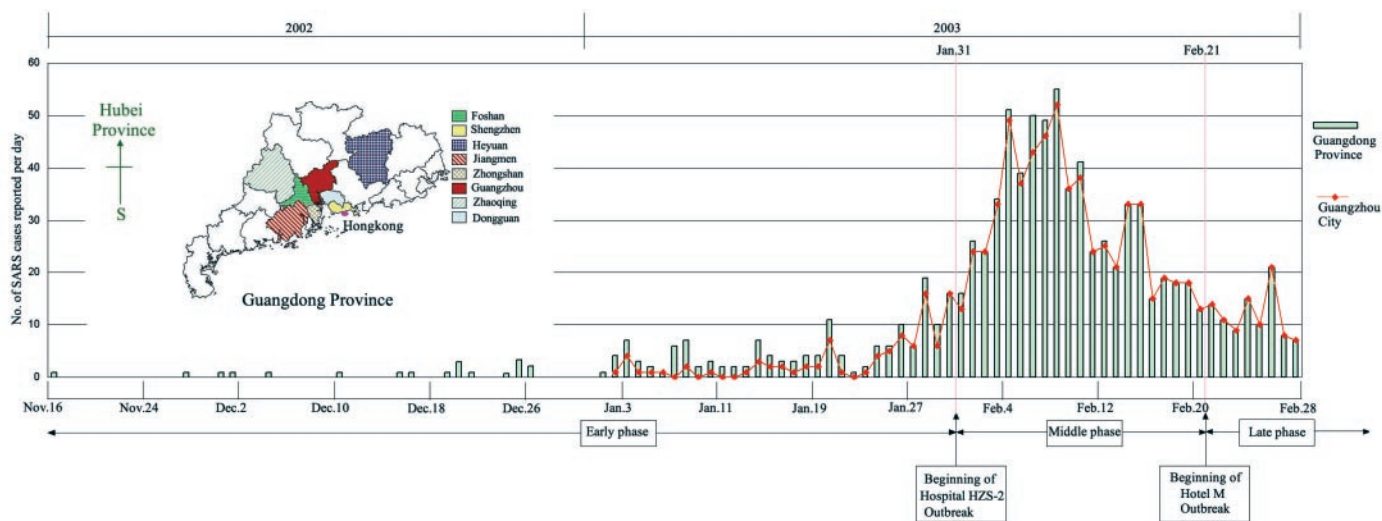


Fig. 1. The triphasic SARS epidemic in Guangdong Province, China. Shown are daily numbers of SARS cases reported in Guangdong Province, in particular the city of Guangzhou. The early, middle, and late phases of the

epidemic are defined in the text. The map shows the geographical distribution of cases belonging to the early phase by administrative districts of Guangdong Province. The detailed data for individual cities are presented in fig. S1.

ological grouping of the genotypes throughout the epidemic (Fig. 2) (table S2).

In tracing the molecular evolution of SARS-CoV in China, we observed that the epidemic started and ended with deletion events, together with a progressive slowing of the nonsynonymous mutation rates and a common genotype that predominated during the latter part of the epidemic. The mechanistic explanation for the selective adaptation and purification processes that led to such genomic evolutionary changes in SARS-CoV requires further work (29). Nonetheless, this study has provided valuable clues to aid further investigation of this remarkable evolutionary tale.

We have sequenced the complete S gene (GenBank accession number AY525636) from an oropharyngeal swab sample (sampling date, 22 December 2003) collected from the most recent index patient of the city of Guangzhou (onset date, 16 December 2003; hospitalized 20 December 2003; www.wpro.who.int/sars/docs/pressreleases/pr_27122003.asp). Phylogenetic analysis of this S gene sequence with those from the human SARS-CoV and palm civet SARS-like coronavirus indicated that this most recent case of SARS-CoV is much closer to the palm civet SARS-like coronavirus than to any human SARS-CoV detected in the previous epidemic (fig. S7 and table S4). Because it is evidently different from the recent laboratory infections in Singapore (www.who.int/csr/don/2003_09_24/en) and Taiwan (www.who.int/mediacentre/releases/2003/np26/en), it strengthens the argument for animal origin of the human SARS epidemic.

References and Notes

1. R. A. Fouchier et al., *Nature* **423**, 240 (2003).
2. T. G. Ksiazek et al., *N. Engl. J. Med.* **348**, 1953 (2003).
3. C. Drosten et al., *N. Engl. J. Med.* **348**, 1967 (2003).
4. P. A. Rota et al., *Science* **300**, 1394 (2003).
5. M. A. Marra et al., *Science* **300**, 1399 (2003).
6. G. Vogel, *Science* **300**, 1062 (2003).
7. Y. J. Ruan et al., *Lancet* **361**, 1779 (2003).
8. Y. Guan et al., *Science* **302**, 276 (2003).
9. S. K. W. Tsui, S. S. C. Chim, Y. M. D. Lo, *N. Engl. J. Med.* **349**, 187 (2003).
10. S. S. C. Chim et al., *Lancet* **362**, 1807 (2003).
11. R. W. K. Chiu, S. S. C. Chim, Y. M. D. Lo, *N. Engl. J. Med.* **349**, 1875 (2003).
12. J. S. Rest, D. P. Mindell, *Infect. Genet. Evol.* **3**, 219 (2003).
13. N. S. Zhong et al., *Lancet* **362**, 1353 (2003).
14. Supporting materials are available on Science Online.
15. K. W. Tsang et al., *N. Engl. J. Med.* **348**, 1977 (2003).
16. N. Lee et al., *N. Engl. J. Med.* **348**, 1986 (2003).
17. Centers for Disease Control and Prevention, *Morb. Mortal. Wkly. Rep.* **52**, 241 (2003).
18. E. J. Snijder et al., *J. Mol. Biol.* **331**, 991 (2003).
19. SARS-like coronaviruses were isolated from palm civets farmed domestically in Hubei Province, China, by Hu et al. at the Wuhan Institute of Virology, Chinese Academy of Sciences. Partial genome sequencing revealed an 82-nt deletion within the Orf8 region, which is identical to that found in human SARS-CoV isolates from the early patients of Zhongshan, Guangdong Province, China. Contamination can be ruled out because no human SARS-CoV isolate with the 82-nt deletion has ever been found in that institute or has been isolated in that region of China.
20. The SARS-CoV sequence with the 415-nt deletion

- (CUHK-LC2, CUHK-LC3, CUHK-LC4, and CUHK-LC5) was obtained from two SARS patients whose disease was linked to a late cluster of SARS cases in Hong Kong. Both patients had disease onset in mid-May 2003. The CUHK-LC2 sequence was initially obtained from the culture isolate of a throat wash specimen of an infected hospital health care worker and was later confirmed from the same specimen directly. CUHK-LC3, CUHK-LC4, and CUHK-LC5 were obtained from three different nasal swab specimens both directly and from the culture supernatants of an elderly patient who acquired SARS in the same hospital.
21. M. M. C. Lai, K. V. Holmes, in *Fields Virology*, D. M. Knipe, P. M. Howley, Eds. (Lippincott Williams & Wilkins, New York, ed. 4, 2001), chap. 35.
22. E. G. Brown, H. Liu, L. C. Kit, S. Baird, M. Nesrallah, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6883 (2001).
23. S. H. Seo, E. Hoffmann, R. G. Webster, *Nature Med.* **8**, 950 (2002).
24. D. Rasschaert, M. Duarte, H. Laude, *J. Gen. Virol.* **71**, 2599 (1990).
25. W.-H. Li, M. Tanimura, P. M. Sharp, *Mol. Biol. Evol.* **5**, 313 (1988).

26. J. W. Drake, J. J. Holland, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13910 (1999).
27. Z. Luo, A. M. Matthews, S. R. Weiss, *J. Virol.* **73**, 8152 (1999).
28. R. M. Bush, C. B. Smith, N. J. Cox, W. M. Fitch, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6974 (2000).
29. P. W. Ewald, *J. Urban Health* **75**, 480 (1998).
30. See SOM Text at Science Online for acknowledgments.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1092002/DC1
 Materials and Methods
 SOM Text
 References and Notes
 Figs. S1 to S7
 Tables S1 to S4

29 September 2003; accepted 14 January 2004
 Published online 29 January 2004;
 10.1126/science.1092002
 Include this information when citing this paper.

Evidence of a Pluripotent Human Embryonic Stem Cell Line Derived from a Cloned Blastocyst

Woo Suk Hwang,^{1,2*} Young June Ryu,¹ Jong Hyuk Park,³
 Eul Soon Park,¹ Eu Gene Lee,¹ Ja Min Koo,⁴ Hyun Yong Jeon,¹
 Byeong Chun Lee,¹ Sung Keun Kang,¹ Sun Jong Kim,³ Curie Ahn,⁵
 Jung Hye Hwang,⁶ Ky Young Park,⁷ Jose B. Cibelli,⁸
 Shin Yong Moon^{5*}

Somatic cell nuclear transfer (SCNT) technology has recently been used to generate animals with a common genetic composition. In this study, we report the derivation of a pluripotent embryonic stem (ES) cell line (SCNT-hES-1) from a cloned human blastocyst. The SCNT-hES-1 cells displayed typical ES cell morphology and cell surface markers and were capable of differentiating into embryoid bodies in vitro and of forming teratomas in vivo containing cell derivatives from all three embryonic germ layers in severe combined immunodeficient mice. After continuous proliferation for more than 70 passages, SCNT-hES-1 cells maintained normal karyotypes and were genetically identical to the somatic nuclear donor cells. Although we cannot completely exclude the possibility that the cells had a parthenogenetic origin, imprinting analyses support a SCNT origin of the derived human ES cells.

The isolation of pluripotent human embryonic stem (ES) cells (1) and breakthroughs in somatic cell nuclear transfer (SCNT) in mammals (2) have raised the possibility of performing human SCNT to generate potentially unlimited sources of undifferentiated

ated cells for use in research, with potential applications in tissue repair and transplantation medicine. This concept, known as “therapeutic cloning,” refers to the transfer of the nucleus of a somatic cell into an enucleated donor oocyte (3). In theory, the oocyte’s cytoplasm would reprogram the transferred nucleus by silencing all the somatic cell genes and activating the embryonic ones. ES cells would be isolated from the inner cell mass (ICM) of the cloned preimplantation embryo. When applied in a therapeutic setting, these cells would carry the nuclear genome of the patient; therefore, it is proposed that after directed cell differentiation, the cells could be transplanted without immune rejection to treat degenerative disorders such as diabetes, osteoarthritis, and Parkinson’s disease

¹College of Veterinary Medicine, ²School of Agricultural Biotechnology, Seoul National University, Seoul 151-742, Korea. ³Medical Research Center, MizMedi Hospital, Seoul, 135-280, Korea. ⁴Gachon Medical School, Incheon, 417-840, Korea. ⁵College of Medicine, Seoul National University, Seoul, 110-744, Korea. ⁶School of Medicine, Hanyang University, Seoul, 471-701, Korea. ⁷College of Natural Science, Suncheon National University, Suncheon, 540-742, Korea. ⁸Department of Animal Science-Physiology, Michigan State University, East Lansing, MI 48824, USA.

*To whom correspondence should be addressed. E-mail: hwangws@snu.ac.kr (W.S.H.); shmoo@plaza.snu.ac.kr (S.Y.M.)

Downloaded from www.sciencemag.org on April 9, 2007