Mice and Oocytes and Zygotes Analyses

Wild-type C57/B6 mice were from the Zhejiang Academy of Medical Science, China. *Vprbp*$^{flox/flox}$ (McCall et al., 2008) and *Zp3-Cre* (Lan et al., 2004) were previously generated. Animal care and experimental procedures were in accordance with the Animal Research Committee guidelines of Zhejiang University. For superovulation assay, pubertal mice (21–23 days old) were injected intraperitonally (i.p.) with 5 IU of pregnant mare serum gonadotrophin (PMSG). After 44 hr, the mice were i.p. injected with 5 IU human chorionic gonadotropin (hCG). After an additional 16 hr, oocyte/cumulus masses were surgically removed from oviducts, and the numbers of oocytes were counted. To obtain fertilized eggs (zygotes), female mice were mated. Zygotes were harvested from oviducts at 20 hr after hCG injection and used for confocal microscopy and real-time PCR analyses.

Electrophoretic Mobility Shift Assay

Biotin-labeled single-stranded oligonucleotides used for TET-DNA binding assay were synthesized and annealed. Immunopurified FLAG-TET2 (500 ng) was incubated in a total 20 μl reaction mixture containing 1 ng of biotin-labeled oligonucleotide. The resulting protein-DNA complexes were resolved in a 6% DNA retardation gel in 0.5 × TBE buffer, transferred to a nylon membrane, and blotted with streptavidin-HRP. The gel-shift assays were performed using the LightShift Chemiluminescent EMSA Kit.

Cell Fractionation

4 × 10^7 U2OS cells were harvested and lysed by Triton X-100. Nuclei were collected in pellet 1 (fraction P1) by low-speed centrifugation (5 min, 1,300 g, 4°C). The supernatant (S1) was further clarified by high-speed centrifugation (10 min, 15,000 g, 4°C) to remove cell debris and insoluble aggregates (P2). Nuclei were washed once in buffer A and then lysed. Insoluble chromatin (P3) was collected by centrifugation (5 min, 1,700 g, 4°C), washed once in buffer B, and centrifuged again under the same conditions. The resulting six fractions were resuspended in the Laemmli buffer, boiled for 15 min, and resolved by SDS-PAGE for western analysis.

In Vitro TET2 Activity Assay

In vitro assay of TET2 catalytic activity was performed using two different substrates, methylated dsDNA oligonucleotides and total genomic DNA. To prepare the TET enzyme, FLAG-tagged wild-type and AML-derived full-length TET2 mutants were ectopically expressed in HEK293T cells, immunoprecipitated, and eluted by FLAG peptide. The reactions were carried out as described in detail in the Supplemental Information, denatured, and then neutralized. The reaction mixtures were spotted on nitrocellulose membrane, crosslinked, and incubated overnight with antibodies recognizing 5mC, 5hmC antibody, 5fC antibody, or 5caC, followed by chemiluminescence.

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures, four tables, and Supplemental Experimental Procedures and can be found with this article at http://dx.doi.org/10.1016/j.molcel.2014.12.002.

ACKNOWLEDGMENTS

We are grateful to Matthew Smith, Kun-Liang Guan, and Yi Zhang for the support and discussion throughout this study; Guoliang Xu for providing Tet3 antibody; Naoya Sasaki for characterizing the VprBP antibody; Shinsuke Ito and Li Shen for technical assistance; Brian Strahl for helping with the synthesis of branched TET2 antigen peptide; and Sarah Jackson, Bryce Seifert, and Howard Fried for reading the manuscript. This study was supported by a Samuel Waxman grant and NIH grants GM067113 and CA163834 to Y.X.

Received: February 10, 2014
Revised: August 26, 2014
Accepted: November 24, 2014
Published: December 31, 2014

REFERENCES

Chowdhury, R., Yeoh, K.K., Tian, Y.M., Hillringhaus, L., Bagg, E.A., Rose, N.R., Leung, I.K., Li, X.S., Woon, E.C., Yang, M., et al. (2011). The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. EMBO Rep. *12*, 463–469.

Costa, Y., Ding, J., Theunissen, T.W., Faiola, F., Hore, T.A., Shliaha, P.V., Fidalgo, M., Saunders, A., Lawrence, M., Dietmann, S., et al. (2013). NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. Nature *495*, 370–374.

Dang, L., White, D.W., Gross, S., Bennett, B.D., Bittinger, M.A., Driggers, E.M., Fantin, V.R., Jang, H.G., Jin, S., Keenan, M.C., et al. (2009). Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. Nature *462*, 739–744.

Delhommeau, F., Dupont, S., Della Valle, V., James, C., Trannoy, S., Massé, A., Kosmider, O., Le Couedic, J.P., Robert, F., Alberdi, A., et al. (2009). Mutation in TET2 in myeloid cancers. N. Engl. J. Med. *360*, 2289–2301.

Deplus, R., Delatte, B., Schwinn, M.K., Defrance, M., Méndez, J., Murphy, N., Dawson, M.A., Volkmar, M., Putmans, P., Calonne, E., et al. (2013). TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. EMBO J. *32*, 645–655.

Doege, C.A., Inoue, K., Yamashita, T., Rhee, D.B., Travis, S., Fujita, R., Guarnieri, P., Bhagat, G., Vanti, W.B., Shih, A., et al. (2012). Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. Nature *488*, 652–655.

Gelsi-Boyer, V., Trouplin, V., Adélaïde, J., Bonansea, J., Cervera, N., Carbuccia, N., Lagarde, A., Prebet, T., Nezri, M., Sainty, D., et al. (2009). Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia. Br. J. Haematol. *145*, 788–800.

Glickman, M.H., and Ciechanover, A. (2002). The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. Physiol. Rev. *82*, 373–428.

Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. Annu. Rev. Biochem. *74*, 481–514.

Gu, T.P., Guo, F., Yang, H., Wu, H.P., Xu, G.F., Liu, W., Xie, Z.G., Shi, L., He, X., Jin, S.G., et al. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature *477*, 606–610.

He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science *333*, 1303–1307.

Hu, L., Li, Z., Cheng, J., Rao, Q., Gong, W., Liu, M., Shi, Y.G., Zhu, J., Wang, P., and Xu, Y. (2013). Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. Cell *155*, 1545–1555.

Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature *466*, 1129–1133.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science *333*, 1300–1303.

Jackson, S., and Xiong, Y. (2009). CRL4s: the CUL4-RING E3 ubiquitin ligases. Trends Biochem. Sci. *34*, 562–570.

Jankowska, A.M., Szpurka, H., Tiu, R.V., Makishima, H., Afable, M., Huh, J., O'Keefe, C.L., Ganetzky, R., McDevitt, M.A., and Maciejewski, J.P. (2009).

(D) Activity of TET2 leukemia mutants was examined by immunofluorescence. See Figure S5D for the dot-blot assay and also Figures S5E and S5F for the similar assays of Tet1 and Tet3 mutants. Also see Figures S5G and S5H for the in vitro assay of the catalytic activity of WT and tumor-derived TET2 mutants.
(E) Summary of ubiquitylation level, VprBP binding, and in vivo and in vitro activity of TET2 leukemia mutants.
See also Figure S5 and Table S2.

Loss of heterozygosity 4q24 and TET2 mutations associated with myelodysplastic/myeloproliferative neoplasms. Blood *113*, 6403–6410.

Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R., et al. (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature *468*, 839–843.

Ko, M., An, J., Bandukwala, H.S., Chavez, L., Aijo, T., Pastor, W.A., Segal, M.F., Li, H., Koh, K.P., Lahdesmaki, H., et al. (2013). Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. Nature *497*, 122–126.

Kosmider, O., Gelsi-Boyer, V., Cheok, M., Grabar, S., Della-Valle, V., Picard, F., Viguié, F., Quesnel, B., Beyne-Rauzy, O., Solary, E., et al.; Groupe Francophone des Myélodysplasies (2009). TET2 mutation is an independent favorable prognostic factor in myelodysplastic syndromes (MDSs). Blood *114*, 3285–3291.

Lan, Z.J., Xu, X., and Cooney, A.J. (2004). Differential oocyte-specific expression of Cre recombinase activity in GDF-9-iCre, Zp3cre, and Msx2Cre transgenic mice. Biol. Reprod. *71*, 1469–1474.

Lewandoski, M., Wassarman, K.M., and Martin, G.R. (1997). Zp3-cre, a transgenic mouse line for the activation or inactivation of loxP-flanked target genes specifically in the female germ line. Curr. Biol. *7*, 148–151.

Li, E. (2002). Chromatin modification and epigenetic reprogramming in mammalian development. Nat. Rev. Genet. *3*, 662–673.

McCall, C.M., Miliani de Marval, P.L., Chastain, P.D., 2nd, Jackson, S.C., He, Y.J., Kotake, Y., Cook, J.G., and Xiong, Y. (2008). Human immunodeficiency virus type 1 Vpr-binding protein VprBP, a WD40 protein associated with the DDB1-CUL4 E3 ubiquitin ligase, is essential for DNA replication and embryonic development. Mol. Cell. Biol. *28*, 5621–5633.

Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S., and Heintz, N. (2012). MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. Cell *151*, 1417–1430.

Nakagawa, T., and Xiong, Y. (2011). X-linked mental retardation gene CUL4B targets ubiquitylation of H3K4 methyltransferase component WDR5 and regulates neuronal gene expression. Mol. Cell *43*, 381–391.

Nakagawa, T., Mondal, K., and Swanson, P.C. (2013). VprBP (DCAF1): a promiscuous substrate recognition subunit that incorporates into both RING-family CRL4 and HECT-family EDD/UBR5 E3 ubiquitin ligases. BMC Mol. Biol. *14*, 22.

Piccolo, F.M., Bagci, H., Brown, K.E., Landeira, D., Soza-Ried, J., Feytout, A., Mooijman, D., Hajkova, P., Leitch, H.G., Tada, T., et al. (2013). Different roles for Tet1 and Tet2 proteins in reprogramming-mediated erasure of imprints induced by EGC fusion. Mol. Cell *49*, 1023–1033.

Quivoron, C., Couronné, L., Della Valle, V., Lopez, C.K., Plo, I., Wagner-Ballon, O., Do Cruzeiro, M., Delhommeau, F., Arnulf, B., Stern, M.H., et al. (2011). TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. Cancer Cell *20*, 25–38.

Rocquain, J., Carbuccia, N., Trouplin, V., Raynaud, S., Murati, A., Nezri, M., Tadrist, Z., Olschwang, S., Vey, N., Birnbaum, D., et al. (2010). Combined mutations of ASXL1, CBL, FLT3, IDH1, IDH2, JAK2, KRAS, NPM1, NRAS, RUNX1, TET2 and WT1 genes in myelodysplastic syndromes and acute myeloid leukemias. BMC Cancer *10*, 401.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930–935.

Tefferi, A., Pardanani, A., Lim, K.H., Abdel-Wahab, O., Lasho, T.L., Patel, J., Gangat, N., Finke, C.M., Schwager, S., Mullally, A., et al. (2009). TET2 mutations and their clinical correlates in polycythemia vera, essential thrombocythemia and myelofibrosis. Leukemia *23*, 905–911.

Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C.J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W., and Walter, J. (2011). 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat. Commun. *2*, 241.

Wu, H., and Zhang, Y. (2014). Reversing DNA methylation: mechanisms, genomics, and biological functions. Cell *156*, 45–68.

Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S.H., Ito, S., Yang, C., Wang, P., Xiao, M.T., et al. (2011). Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α-ketoglutarate-dependent dioxygenases. Cancer Cell *19*, 17–30.

Xu, Y., Xu, C., Kato, A., Tempel, W., Abreu, J.G., Bian, C., Hu, Y., Hu, D., Zhao, B., Cerovina, T., et al. (2012). Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development. Cell *151*, 1200–1213.

Yamaguchi, S., Hong, K., Liu, R., Shen, L., Inoue, A., Diep, D., Zhang, K., and Zhang, Y. (2012). Tet1 controls meiosis by regulating meiotic gene expression. Nature *492*, 443–447.

Yang, H., Liu, Y., Bai, F., Zhang, J.Y., Ma, S.H., Liu, J., Xu, Z.D., Zhu, H.G., Ling, Z.Q., Ye, D., et al. (2013). Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. Oncogene *32*, 663–669.

Yildirim, O., Li, R., Hung, J.H., Chen, P.B., Dong, X., Ee, L.S., Weng, Z., Rando, O.J., and Fazzio, T.G. (2011). Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. Cell *147*, 1498–1510.

短　報

舞踏運動を呈したdysferlin異常症の1例

髙橋　俊明[1)]　今井　尚志[1)]　田中　洋康[1)]　吉岡　勝[1)]
今野　秀彦[1)]　日向野修一[2)]　小野寺好明[3)]　斎藤　博[1)]
木村　格[1)]　糸山　泰人[3)]　武田　篤[1)]　青木　正志[3)]

要旨：Dysferlin異常症は常染色体劣性遺伝形式をとり，三好型遠位型筋ジストロフィーや肢帯型筋ジストロフィー2B型を主な表現型とする．症例は53歳男性．31歳に筋力低下で発症し，37歳から不随意運動が徐々に進行した．左優位に近位筋力低下，病的反射，排尿障害，舞踏運動が認められた．Dysferlin遺伝子にc.2997G>T変異がホモ接合で認められた．舞踏運動をきたす他の疾患の合併は見出せなかった．Dysferlinの脳における働きやdysferlin異常症の中枢神経系の研究はほとんど行われていない．今後この分野の研究も必要と思われた．

(JMDD 2014; 24: 51–54)

緒　言

Dysferlin異常症は2番染色体に位置するdysferlin遺伝子の変異を原因とする．常染色体劣性遺伝形式をとり，三好型遠位型筋ジストロフィーや肢帯型筋ジストロフィー2B型を主な表現型とする．我々は筋ジストロフィーでありながら舞踏運動を呈したdysferlin異常症の1症例を経験し，2006年に報告[1)]した．今回，その後の研究報告の結果を考察に加え報告する．

症例提示

53歳男性．両親が血族婚．神経筋疾患や不随意運動の家族歴はない．31歳時に立ち上がる際の筋力低下に気付いた．37歳で肢帯型筋ジストロフィーと診断．同年妻が，症例の手が持続的に動いていることに気付く．筋力低下が徐々に進行，不随意運動は四肢に広がった．50歳時には本人も不随意運動が気になるようになった．49歳時から頻尿のためプロピベリン塩酸塩を服用していたが，1カ月の休薬で不随意運動は改善しなかった．

認知機能に問題なし．やや速めの話し方．著明な四肢近位優位の筋力低下と筋萎縮．肩の筋力が著明に低下し，頸，肘の屈曲，膝の伸展および足の底屈が次いで障害されていた．腸腰筋と大腿四頭筋は左がより弱かった．反射は上腕三頭筋と尺骨以外消失．伸展性足底反応でRossolimo反射は左が陽性．左優位に頸，肩と四肢に舞踏運動が見られた．計算や会話で増強した．

Key words：dysferlin，肢帯型筋ジストロフィー2B型，舞踏運動

1) 独立行政法人国立病院機構仙台西多賀病院神経内科
(〒982-8555　仙台市太白区鈎取本町2丁目11番11号　TEL 022-245-2111)
2) 東北大学大学院医学系研究科量子診断学
3) 東北大学大学院医学系研究科神経内科

(受理日　2014年6月16日)

図1 症例の頭部MRIのT2強調像．橋，両視床，右尾状核頭近傍に高信号領域が認められた．くも膜嚢胞も小脳背側に見られた．（著作権の関係から文献1と異なる写真である．）

Mini-Mental State試験は28点，Wechsler成人知能検査改定版でIQ86．頭部MRIで軽度の所見が認められた（図1）．脊髄MRIは正常．CKが1,837 IU/l．髄液，末梢血像，赤沈，CRP，血清銅，セルロプラスミン，抗核抗体，抗リン脂質抗体，ASO，ASK，甲状腺ホルモンは正常．Dysferlin遺伝子にc.2997G>T変異がホモ接合で認められた．Huntington病，歯状核赤核淡蒼球ルイ体萎縮症の原因遺伝子は正常だった．

考 察

2006年の本症例報告後も，検索し得た範囲でdysferlin異常症が舞踏運動を呈した報告はない．我々は最近日本人40例の肢帯型筋ジストロフィー2B型の臨床的特徴を報告したが，その中にも舞踏運動を示したものは本例のみだった[2]．我々の検索で有棘赤血球舞踏病，Wilson病，全身性エリトマトーデス，甲状腺機能亢進症，Huntington病，歯状核赤核淡蒼球ルイ体萎縮症，プロピベリン塩酸塩の副作用，脳梗塞などは否定できる[1]ものと考えた．しかし舞踏運動を呈する疾患は他にも多数あり，今後も他の疾患の合併の徴候がないかは注意深く観察していく必要がある．

Dysferlin異常症に中枢神経合併症を呈した報告は，2006年以降2件見出すことができた．山中らは若年性脳梗塞をきたした39歳女性と脳梗塞とWillis動脈輪閉塞症を合併した42歳女性の姉妹例[3]を，和田らは認知症を併発した3同胞例のdysferlin異常症[4]を報告している．2006年時点では，dysferlin異常症の頭部MRI所見の報告は見出せなかった．上記の山中らの報告の39歳女性[3]と和田らの報告の発端者[4]では，合併症に見合う頭部MRI所見が報告されている．今回，自験例数例ではあるが，他のdysferlin異常症例の頭部MRIを見直してみたが，異常所見を見つけることはできなかった．なお，この他の頭部MRI所見の報告を見出すことはできなかった．

本症例報告の2006年，dysferlinは中枢神経系の錐体細胞や扁桃体の投射ニューロンの細胞質に局在する[5]との報告がみられた．しかし，dysferlinの脳における働きやdysferlin異常症の中枢神経系の研究はほとんど行われていない．今後この分野の研究も必要と思われた．

謝 辞

本研究にご協力いただきました患者さま，dysferlin遺伝子解析を手伝っていただいた国立療養所西多賀病院臨床検査科の阿部恵美さんおよび助言をいただいた国立病院機構仙台西多賀病院神経内科の谷口さやか先生に深謝いたします．本研究は国立精神・神経医療研究センター精神・神経疾患研究開発費（23－4）および（26－7）によって行われた．

文 献

1) Takahashi T, Aoki M, Imai T, et al. A case of dysferlinopathy presenting choreic movements. Mov Disord 2006; 21: 1513–1515.
2) Takahashi T, Aoki M, Suzuki N, et al. Clinical features and a mutation with late onset of limb girdle muscular dystrophy 2B. J Neurol Neurosurg Psychiatry 2013; 84: 433–440.
3) 山中敏之，松井聖博，橋本満喜子ら．若年で脳血管障害を合併したDysferlin欠損症の姉妹例．臨床神経 2012; 52: 130.
4) 和田千鶴，小原講二，阿部エリカら．認知症を伴った肢帯型筋ジストロフィー2B型の1家系．臨床神経 2013; 53: 754.
5) Galvin JE, Palamand D, Strider J, et al. The muscle protein dysferlin accumulates in the Alzheimer brain. Acta Neuropathol 2006; 112: 665–671.

Abstract

A case of dysferlinopathy presenting choreic movements

Toshiaki Takahashi[1)], Takashi Imai[1)], Hiroyasu Tanaka[1)], Masaru Yoshioka[1)], Hidehiko Konno[1)], Shuichi Higano[2)], Yoshiaki Onodera[3)], Hiroshi Saito[1)], Itaru Kimura[1)], Yasuto Itoyama[3)], Atsushi Takeda[1)] and Masashi Aoki[3)]

[1)]Department of Neurology and Division of Clinical Research, National Hospital Organization Sendai-Nishitaga National Hospital
[2)]Department of Radiology, Tohoku University School of Medicine
[3)]Department of Neurology, Tohoku University School of Medicine

Mutations in the dysferlin gene cause limb-girdle muscular dystrophy type 2B. We report the case of a 53-year-old man who had progressive muscle weakness and involuntary movements. He first noticed weakness at the age of 31. His wife first noticed involuntary movements when he was 37. The weakness and movements progressed slowly. Neurological examination revealed muscle weakness and atrophy in the limb-girdle, extensor plantar response, pollakisuria, and choreiform movements. The involuntary movements increased during mental tasks and speaking, and occured in the neck, the shoulders and all four limbs. A homozygous c.2997G>T mutation was identified in the dysferlin gene. The patient had no evidence of other causes of chorea. There have been few reports of the function of the isoform of the dysferlin gene, or of brain imaging or symptoms of the central nervous system in patients with dysferlinopathy. Further study is needed to determine the significance of the isoform of dysferlin in the brain and the involvement of the central nervous system in dysferlinopathy.

(JMDD 2014; 24: 51–54)

Key words: dysferlin, limb-girdle muscular dystrophy type 2B, chorea

Journal of Human Genetics (2014) 59, 326–331
© 2014 The Japan Society of Human Genetics All rights reserved 1434-5161/14
www.nature.com/jhg

ORIGINAL ARTICLE

Compilation of copy number variants identified in phenotypically normal and parous Japanese women

Ohsuke Migita[1,11], Kayoko Maehara[1,11], Hiromi Kamura[1], Kei Miyakoshi[2], Mamoru Tanaka[3], Seiichi Morokuma[4], Kotaro Fukushima[4], Tomihiro Shimamoto[5], Shigeru Saito[6], Haruhiko Sago[7], Keiichiro Nishihama[8], Kosei Abe[1], Kazuhiko Nakabayashi[1], Akihiro Umezawa[9], Kohji Okamura[10] and Kenichiro Hata[1]

With increasing public concern about infertility and the frequent involvement of chromosomal anomalies in miscarriage, analyses of copy number variations (CNVs) have been used to identify the genomic regions responsible for each process of childbearing. Although associations between CNVs and diseases have been reported, many CNVs have also been identified in healthy individuals. Like other types of mutations, phenotypically indefinite CNVs may have been retained and accumulated during anthropogenesis. Therefore to distinguish causative variants from other variants is a formidable task. Furthermore, because previous studies have predominantly focused on European and African populations, comprehensive detection of common Asian CNVs is eagerly awaited. Here, using a high-resolution genotyping array and samples from 411 Japanese women with normal parity without significant complications, we have compiled 1043 copy number variable regions. In total, the collected regions cover 164 Mb, or up to 0.5% of the genome. The copy number differences in these regions may be irrelevant not only to infertility but also to a wide range of diseases. The utility of this resource in reducing the candidate pathogenetic variants, especially in Japanese subjects, is also demonstrated.
Journal of Human Genetics (2014) **59,** 326–331; doi:10.1038/jhg.2014.27; published online 1 May 2014

INTRODUCTION

The advent of new technologies has allowed the identification of structural variants that have a more significant impact on human diversity than does the entire set of single-nucleotide polymorphisms (SNPs). Copy number variations (CNVs) are one such type of structural variant and constitute the largest proportion of genomic variations.[1–3] CNVs result from the duplication or deletion of a DNA segment and are commonly observed in human genomes.[4–7] When a genomic event results in a CNV, not only the copy number of a gene can be altered but also its genic sequences. Therefore, CNVs can cause disease or contribute to disease susceptibility,[8–10] and they have been compiled in several databases for public use.[9–11]

Although a number of deleterious changes may have been negatively selected during human evolution, it is likely that phenotypically neutral changes have been retained, transmitted and accumulated over generations. Increasing numbers of CNVs are found in phenotypically normal human individuals.[1] Accordingly, each ethnic group tends to have distinct features in terms of the positions, copy numbers and frequencies of their CNVs, and it is possible that fixed CNVs have contributed to ethnic differences in phenotypic variations and disease susceptibility.[12–15] Therefore, it is important to have a list of CNVs for each ethnic group, especially for medical purposes. However, the number of reported CNVs from Asian populations is small compared with those of Europeans and Africans. Extensive examination of Asian CNVs is eagerly awaited by Asian researchers.

The compilation of nonpathogenic variations, in addition to disease-related variations, is also important for a better understanding of the genetic landscape of the human genome. Data sets including both these sorts of variations should be helpful in pinpointing causative mutations. Even when we search for variations using patient samples, most of the variations identified would be normal polymorphisms, together with a few pathogenic mutations. Although we can consider most of the available variation data nonpathogenic, it is difficult to know which variations are pathogenic. Therefore, the collection of data from normal controls is essential. To investigate

[1]Department of Maternal–Fetal Biology, National Research Institute for Child Health and Development, Tokyo, Japan; [2]Department of Obstetrics and Gynecology, School of Medicine, Keio University, Tokyo, Japan; [3]Department of Obstetrics and Gynecology, St Marianna University School of Medicine, Kanagawa, Japan; [4]Department of Obstetrics and Gynecology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan; [5]Department of Obstetrics and Gynecology, Miyazaki Prefectural Miyazaki Hospital, Miyazaki, Japan; [6]Department of Obstetrics and Gynecology, University of Toyama, Toyama, Japan; [7]Department of Maternal–Fetal and Neonatal Medicine, National Center for Child Health and Development, Tokyo, Japan; [8]Illumina KK, Tokyo, Japan; [9]Department of Reproductive Biology and Pathology, National Research Institute for Child Health and Development, Tokyo, Japan and [10]Department of Systems BioMedicine, National Research Institute for Child Health and Development, Tokyo, Japan
[11]These authors contributed equally to this work.
Correspondence: Dr K Okamura or Dr K Hata, Department of Maternal–Fetal Biology, National Research Institute for Child Health and Development, 2-10-1 Okura, Setagaya Ward, Tokyo 157-8535, Japan.
E-mail: okamura-k@ncchd.go.jp or hata-k@ncchd.go.jp
Received 27 September 2013; revised 21 March 2014; accepted 26 March 2014; published online 1 May 2014

phenotypically 'normal' samples in this study, we considered reproduction and child development, and chose parous Japanese women, who had experienced normal pregnancies and deliveries.

Although the origin of the Japanese population remains controversial, the last major migration to the Japanese Archipelago is thought to have occurred approximately 2000 years ago.[16,17] The population has been mixed well with various Asian ethnic groups during previous migrations, but has remained relatively isolated for 2000 years. However, although the current population of Japan is 127 million, far fewer CNVs have been documented in Japanese samples than in Europeans. Compiling a list of Japanese CNVs is also important from the perspective of medical science in Japan.

MATERIALS AND METHODS

Subject recruitment and SNP genotyping with a high-resolution microarray

We examined 411 unrelated Japanese women who had had one or more normal parities, with no significant abnormalities in any of their pregnancies, deliveries or neonates. Ethical approval was also obtained from each review board of the hospitals that participated in the study. The informed consent of all the subjects was obtained. To avoid cell-culture-induced chromosomal rearrangements, genomic DNAs were extracted directly from blood using the QIAsymphony DNA Midi Kit (Qiagen, Venlo, The Netherlands) with the QIAsymphony SP instrument and analyzed with a high-resolution SNP-based genotyping microarray, HumanOmni2.5-8 BeadChip (Illumina, San Diego, CA, USA). Only data that met the quality control guidelines of the manufacturer were used for further analyses.

Identification of CNVs and CNVRs

Two distinct algorithms were used to maximize the specificity of our CNV calling: a likelihood-based method with CNVPartition version 3.2.0 (http://www.illumina.com/software/illumina_connect.ilmn) and a hidden Malkov method with PennCNV version (27 August 2009).[18] The parameters applied with these tools were referred to those typically used by many research groups (at least three consecutive probes to define a CNV, using the GC wave adjustment option, etc.). These programs computed confidence scores that can be used to filter out CNV regions that are likely to be false positives. However, we should note that the two programs calculated the scores in different ways, with different scales. To minimize false positives, we first chose only CNVs with high confidence scores; that is, more than 100 with CNVPartition, and selected copy number variable regions (CNVRs) that overlapped those called by PennCNV for at least 80% of their lengths. For PennCNV, we generated a list of B allele frequencies using a collection of signal intensities for 47 samples from HapMap Japanese in Tokyo with the compile_pfb script (Figure 1a).

Multiplex PCR assay

Multiplex polymerase chain reaction (PCR) assay was used to confirm regions that had been called homozygously deleted. The reactions were performed with both a control primer pair that generated a 296-bp fragment and a test primer pair that amplified a target region. The thermal cycling conditions were initial denaturation at 95 °C for 2 min, followed by 35 cycles of denaturation at 95 °C for 30 s, annealing at 60 °C for 30 s and extension at 72 °C for 30 s, and a final extension at 72 °C for 3 min. Detailed information on these primers is given in Supplementary Table S3.

RESULTS

Genetic ancestry of the subjects

First, the population structure was inferred with the Structure software (http://pritchardlab.stanford.edu/structure.html) to confirm the Japanese ancestry of the subjects.[19] A cluster analysis of our samples together with the sequences of 499 HapMap individuals from three ancestral populations (European, African and Asian) was performed using 1959 unlinked tag SNPs on chromosome 21. The expected ancestry of all the subjects was confirmed with a minimum coefficient of 0.85. We also performed a principal components analysis with the pca.jar program (Biobank Japan project; http://genome-analysis.src.riken.jp/PCP/). The results indicated that all but one subject were derived from the main islands of Japan and that the remaining singleton was Ryukyuan.[20]

Characterization of CNVs and CNVRs

The CNVPartition software (Illumina) identified 26 150 candidate regions as CNVs. We then used another program, PennCNV,[18] which is based on an integrated hidden Markov algorithm, to maximize the specificity of the analysis. If a candidate CNV was also supported by PennCNV for at least 80% of its length, it was retained. In this way,

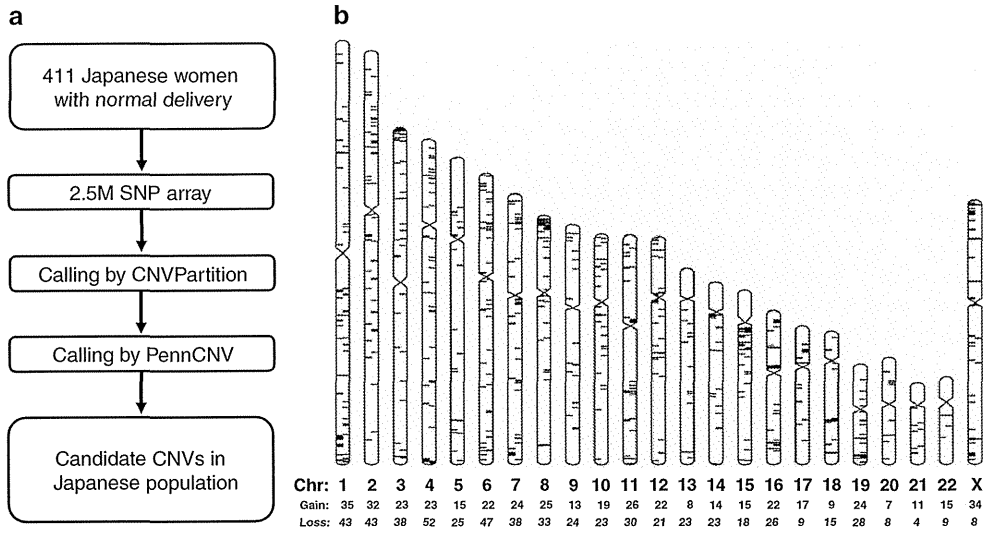


Figure 1 (**a**) Data processing flow. The initial 26 150 regions identified with CNVPartition were validated with PennCNV. (**b**) Chromosomal distribution of the CNVRs. Each CNVR is shown by a horizontal bar. Gain- and loss-type CNVs are distinguished by bars on the left and right, respectively. The numbers of each type of CNV are also shown, and are drawn with Idiographica (http://www.ncrna.org/idiographica/). CNV, copy number variation; CNVR, copy number variable regions.

Table 1 Comparison of the CNVRs with those reported in other studies and in the DGVs

	Present study	*McCarroll et al.*[21]	*Conrad et al.*[22]	*Koike et al.*[23]	*DGV Jul. 2013*
CNVs reported	1043	592	1768	169	202 430
CNVs spanning our data	1043	88 (71/1043)[a]	156 (112/1043)[a]	37 (45/1043)[a]	30 322 (1033/1043)[a]
Number of samples	411 Japanese females	45 HapMap JPT	45 HapMap JPT	57 Japanese females and 123 Japanese males	Collective (including non-Japanese samples)
Experimental method	SNP array (Illumina HumanOmni2.5-8 BeadChip)	SNP array (Affymetrics Genome-Wide Human SNP Array 6.0)	Custom CGH array (NimbleGen and Agilent)	SNP array (Affymetrics Genome-Wide Human SNP Array 6.0)	N/A
CNV calling	CNVPartition and then PennCNV	Birdseye and custom program	Custom program	PennCNV	N/A

Abbreviations: CNV, copy number variations; CNVR, copy number variable region; DGV, database of genomic variants; JPT, Japanese in Tokyo; N/A, not available; SNP, single-nucleotide polymorphism.
[a]The number of CNVRs overlapped with those in the present study is indicated within parentheses.

Table 2 CNVRs overlapping between the Japanese and other populations

Population	*Sample size*	*Reported CNVRs*	*Frequency of overlapping regions among studies*[a]
Japanese (present study)	411	1043	—
Korean[24]	100	576	10% (106/1043)
Tibetan[14]	29	139	4.9% (51/1043)
Chinese[13] (Han, Tibetan and five other ethnic group)	155	1440	17% (173/1043)
Han Chinese[13]	80	1407	17% (175/1043)
Swiss[25]	717	917	16% (163/1043)
Rwandan[25] (sub-Saharan African)	450	1185	14% (141/1043)
HapMap[26] (mixed)	112	3262	13% (134/1043)

Abbreviation: CNVR, copy number variable regions.
[a]Number of overlapped CNVRs is indicated within parentheses.

Table 3 List of genes lying within a homozygously deleted region

No.	*Coordination*	*Frequency*	*Suffered gene*		
1	Chr 1: 161 570 803–161 644 281*	2/411	*FCGR3B*	*FCGR2B*	
2	Chr 2: 111 884 593–111 886 246*	3/411	*BCL2L11*		
3	Chr 4: 69 367 146–69 489 473*	302/411	*UGT2B17*		
4	Chr 5: 180 377 470–180 424 820*	32/411	*BTNL3*		
5	Chr 6: 32 551 892–32 555 728	2/411	*HLA-DRB1*		
6	Chr 7: 115 584 568–115 593 688*	1/411	*TFEC*		
7	Chr 7: 141 761 027–141 795 404*	6/411	*MGAM*		
8	Chr 11: 18 949 220–18 961 743	1/411	*MRGPRX1*		
9	Chr 19: 41 350 895–41 379 321*	10/411	*CYP2A6*		
10	Chr 19: 43 590 229–43 772 302	86/411	*PSG5*	*PSG4*	*PSG9*
11	Chr 19: 46 622 776–46 636 139*	3/411	*IGFL3*		
12	Chr 19: 52 132 392–52 150 601*	114/411	*SIGLEC5*		

Abbreviation: PCR, polymerase chain reaction.
Asterisk indicates a homozygously deleted region validated by PCR.

we identified 6871 CNVs and 1043 regions with variable copy numbers from 411 Japanese individuals, with an average of 16.7 CNVs per diploid genome (Supplementary Table S1). Detailed information on all the SNP probes used for the CNV calls is tabulated (Supplementary Table S2). The mean length of the CNVs was 79.9 kb, ranging from 169 bases to 2.27 Mb. These 6871 CNVs corresponded to 1043 CNVRs (588 losses and 455 gains). Figure 1 shows the chromosomal distribution of the observed CNVRs. The total length of all of these CNVRs was 163 720 kb, which is equivalent to 0.5% of the whole human genome. The CNVRs can be divided into gain regions and loss regions, depending on whether their copy numbers have increased or decreased. Of the 1043 regions identified, 1033 overlap the latest database of genomic variants (DGVs) (released on 23 July 2013) reported at the DGV. More than half the CNVRs, including 72% of the gain CNVRs and 36% of the loss CNVRs, intersect RefSeq gene loci.

As far as we know, three studies have examined the Japanese population with array-based methods: two of them used samples from HapMap and the other used healthy individuals.[21–23] These results are summarized with our data set (Table 1). Although those three studies had together already reported 82 regions, more than half the regions reported in the present study were not detected by them. It is probable that the higher resolution of our analysis and our larger sample size allowed us to detect additional CNVRs. Depending on the types of platform used, array-based CNV studies occasionally show discrepancies in the regions of CNVs.[21,22] Differences in the array architectures, scanning machines and calling algorithms could affect the final data sets. Using reported CNV data from SNP arrays, we counted the overlapping regions among studies that focused on other populations or HapMap data[13,14,24–26] (Table 2). The similarities among these studies are comparable, but our results suggest a greater similarity between the Japanese and Chinese populations.

Homozygous deletions found in parous Japanese women

In our study, 1628 homozygous deletions that could affect 112 RefSeq gene loci were called in a total of 822 chromosomes. Although the CNV analysis was unable to determine the precise breakpoints, our data indicate that some exonic sequences are disrupted by homozygous deletions (Table 3). Using multiplex PCR with both control and test primer pairs, we confirmed the null genotypes caused by deletions (Supplementary Figure S1 and Supplementary Table S3). Five genes, *FCGR3B*, *FCGR2B*, *UGT2B17*, *HLA-DRB1* and *CYP2A6*, are described as disease related in the OMIM database. The *FCGR3*, *FCGR2B* and *HLA-DRB1* genes have roles in the immune system. *FCGR3B* and *FCGR2B* encode the crystallizable region of immunoglobulin G. Several studies have shown that a low copy number at the *FCGR3B–FCGR2B* locus is associated with a susceptibility to systemic lupus erythematosus in the Caucasian population,[27–29] but not in the Chinese population.[28] *UGT2B17* encodes a protein that belongs to the family of UDP-glucuronosyltransferases enzymes, which catalyzes the glucuronidation of steroid hormones. A case–control study of

osteoporosis-related fracture suggested that a CNV at the *UGT2B17* locus contributes to osteoporosis.[30] Jakobsson *et al.*[31] found that its null genotype was more common in Koreans (67%) than in Swedish (9%). Our array results also showed a high frequency (74%) of the null genotype. The CYP2A6 protein metabolizes nicotine and coumarin in the liver. The lack of a *CYP2A6* gene may affect nicotine levels in individuals and probably has a protective effect against tobacco dependence.[32] Another study reported that the frequency of homozygotes for the *CYP2A6* gene deletion was lower in Japanese lung cancer patients than in control samples.[33] Except for *HLA-DRB1*, these disease-related genes have been reported to be frequently deleted in Asian populations.[25,34–36] Because we limited our samples to parous women only, it is unlikely that the CNVRs identified in the present study are related to human reproduction.

DISCUSSION

In the present study, we compiled a catalog of copy number variable regions identified in phenotypically normal Japanese samples, especially those with a history of full-term pregnancy and deliveries without major complications. The data set will be useful in the search for novel or rare CNVs that increase the individual's susceptibility to congenital diseases and complications during pregnancy. It is unlikely that the newly identified CNVs are related to infertility or miscarriage. CNVs in parous women without complications have never before

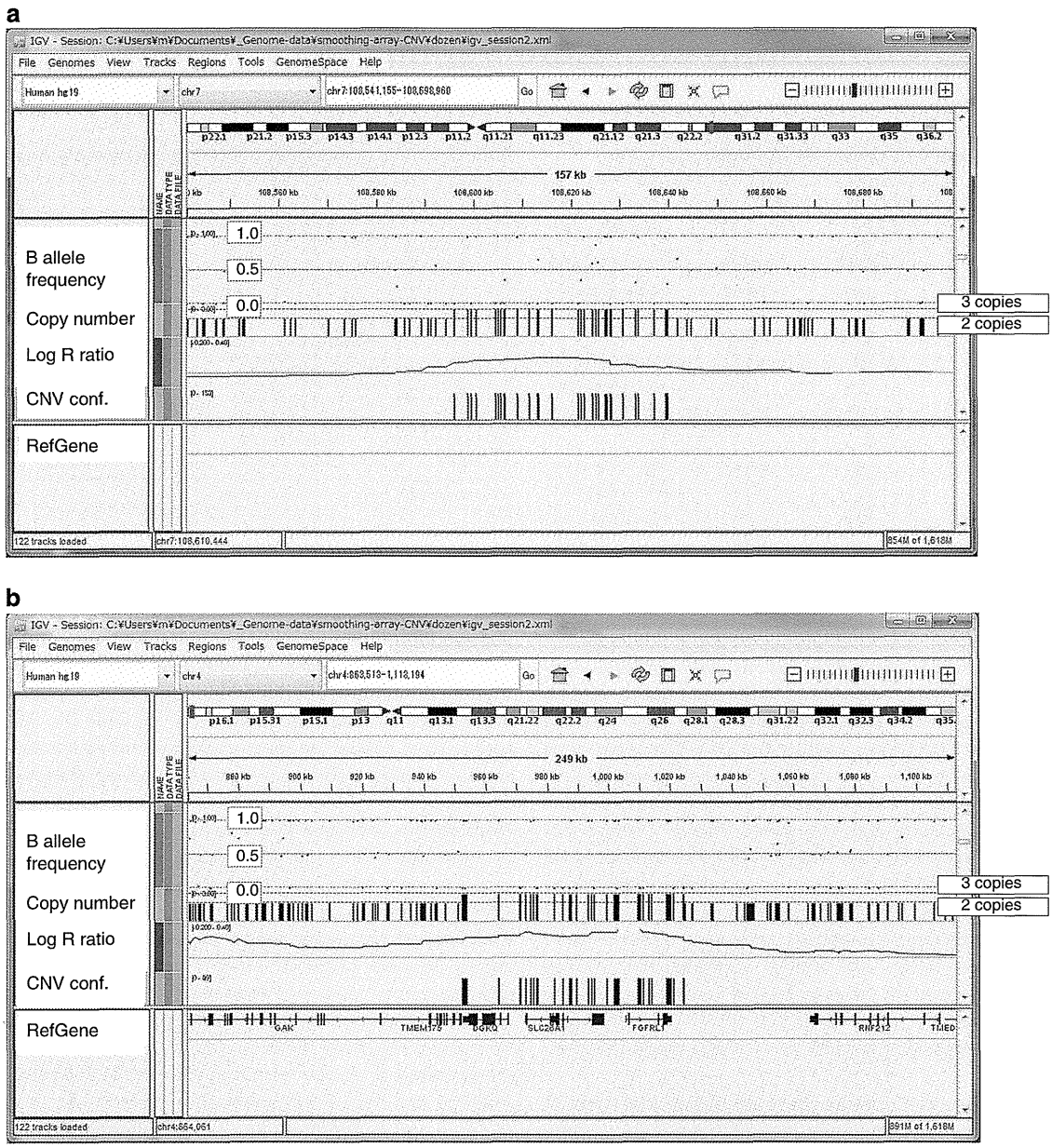


Figure 2 (**a**) A copy number variation (CNV) located on chromosome 7. The panel shows the region at nucleotides 108 541 155–108 698 960 in hg19. This CNV was called with high-intensity probes. The B allele frequencies (BAFs) were separated into four levels, which corresponded to AAA, AAB, ABB and BBB, respectively. (**b**) Another CNV located in the subtelomeric region on chromosome 4. The panel shows the region at nucleotides 863 513–1 113 194. Despite high-intensity probes used, as in the example shown above, the four levels of BAFs were not observed, suggesting that the call might be implausible. Such CNVs tended to be called in G+C-rich regions; for example, 58% G+C content in this case. The snapshot was made with the IGV program (http://www.broadinstitute.org/igv/). A full color version of this figure is available at *Journal of Human Genetics* online.

been investigated. Although the copy numbers of these regions were not thoroughly validated with other methods; such as, quantitative PCR, according to DGV, most of the CNVRs identified here have been reported in previous studies, indicating that they should be observed by other methods or techniques. Because our identification strategy was based on a microarray technique, it is inevitable that errors would have occurred. Besides routine data processing, we also carefully curated the data by examining the B allele frequencies and signal intensities (log *R* ratio) for each CNVR using the GenomeStudio software (Illumina) (Figure 2). We found that many implausible calls were situated in regions with high G + C contents; for example, in subtelomeric regions. All of them were copy number gain-type CNVs rather than copy number loss-type CNVs. Although further research is required, it is important to note that CNVRs tend to be detected in those regions by SNP microarrays. Even if such CNVRs are false positives, our data set is still useful for screening large numbers of candidate CNVs.

It is unclear whether CNVs are selectively neutral on the basis of genetic drift, but they are certainly distributed throughout all human populations. Using the genotypes of mitochondrial DNA and Y chromosome, geneticists and anthropologists have surmised various intriguing scenarios about the history of humans.[37–40] However, these genetic materials have been transmitted exclusively through maternal and paternal lineages, respectively. In contrast, the CNVs reported here occur in the more extensive remaining genome regions; that is, on autosomes or the X chromosome. Therefore, they have acted some times as maternal alleles at and at other times as paternal alleles. They might also have been subjected to crossingover. CNV data from various parts of the world are essential to substantiate these hypothetical scenarios.

Chromosomal anomalies are found with conventional cytogenetic techniques in approximately half of all early sporadic miscarriages.[41] It is possible that miscarriages and pregnancy losses are also caused by submicroscopic chromosomal changes, including CNVs. Twenty-eight CNVs have been reported as candidate miscarriage-related variations when instances of recurrent pregnancy loss were examined by Rajcan-Separovic *et al.*[42] When 17 Caucasian and three African-American couples with recurrent pregnancy losses and their miscarriage samples were examined, CNVs that may have been related to miscarriages were reported.[42] They reported 11 novel CNVs in miscarriage samples and three in the parent samples and suggested that these CNVs were probably mutations causing susceptibility to miscarriage. Of the 11 CNVs in the miscarriage samples, one on chromosome 12 (130 060 706–130 430 847 in hg18) and another one on chromosome X (6 498 521–8 091 951) overlapped with our data set. Whereas the first one on chromosome 12 was up to 370 kb in length and encompassed the *GPR133* gene, the corresponding variable region in our data set is much shorter and includes no known genes. The *GPR133* gene encodes one of the orphan G-protein-coupled receptors, but its function is unknown.[43] It is possible that this receptor protein has a role in several signal-transduction pathways via classical receptor/G-protein interactions. Therefore, the CNV mentioned above may be a variant that causes miscarriage. However, one of the CNVs on chromosome X is consistent with our data set, suggesting that it is a commonly observed variant. In fact, Rajcan-Separovic *et al.*[42] tried to define the common CNVs using a collective repository in the DGV, but insufficient phenotypic information was available to refine the data. Taking these observations together, it seems that to define a set of common CNVs, it will be necessary to collect a large number of control data that focus on a specific phenotype; such as, normal parity in this case.

The Japanese are an admixture of ancient Asian populations that inhabited regions outside the Japanese Archipelago. We investigated the similarities among the CNVRs detected in various populations and noted that around 15% of Japanese CNVRs overlap those of other populations (Table 2). It has been suggested that the number of overlapping CNVs is influenced by the number of subjects. For instance, Japanese and Tibetan data showed dissimilarity because of the limited number of Tibetan subjects. Although the sample sizes of the Korean and Chinese populations are smaller than those of the European and African populations, similarities between the Japanese and other East Asian populations were similar to those of the European and African populations. This probably suggests strong similarities between the Japanese and other East Asian populations.

Previous studies have predominantly targeted European and African populations, but CNVs have been observed at different frequencies or copy numbers in different populations; for example, variations in the salivary amylase gene.[44] Many CNVs; such as, those at the *AMY1* locus, may be associated with diabetes, asthma, hypertension, allergy and other diseases of affluence in each ethnic group. Although CNVRs may result from the accumulation of tolerable structural mutations in the course of an ethnic history, they could start to influence the population's susceptibility to disease once its lifestyle is altered. The allelic frequencies of SNPs and short indels in each population have recently been documented.[45] The complete documentation of the CNVRs in each ethnic group is similarly important. The development of an innovative method to achieve this; such as, one involving next-generation sequencing and informatics, is another challenge.

CONFLICT OF INTEREST

The authors received no financial support from Illumina KK and the company had no role in the study design. The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

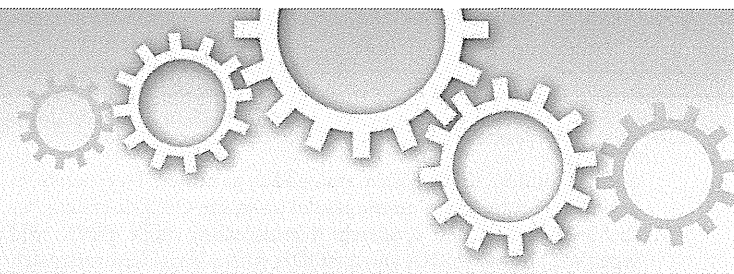
We are grateful to all the participants in the present study, including the 411 women. Computation time was partly provided by the supercomputer system, Shirokane, at the Human Genome Centre, Institute of Medical Science, University of Tokyo. This work was supported by CREST Program 'Epigenomic analysis of the human placenta and endometrium constituting the fetal-maternal interface' of Japan Science and Technology Agency (JST) and Health and Labor Sciences Research Grants for Research into Rare and Intractable Diseases (H23 Jitsuyoka (Nanbyo)-Ippan-003 and H25 Jisedai-Ippan-001), and was also partly supported by KAKENHI 23770273, 24657151, 24390251, 24592494, 24659742, 25293345 and 25860258.

1 Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D. *et al.* Global variation in copy number in the human genome. *Nature* **444,** 444–454 (2006).
2 Stranger, B. E., Forrest, M. S., Dunning, M., Ingle, C. E., Beazley, C., Thorne, N. *et al.* Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315,** 848–853 (2007).
3 Pang, A. W., Migita, O., Macdonald, J. R., Feuk, L. & Scherer, S. W. Mechanisms of formation of structural variation in a fully sequenced human genome. *Hum. Mut.* **34,** 345–354 (2013).
4 Frazer, K. A., Chen, X., Hinds, D. A., Pant, P. V., Patil, N. & Cox, D. R. Genomic DNA insertions and deletions occur frequently between humans and nonhuman primates. *Genome Res.* **13,** 341–346 (2003).
5 Locke, D. P., Segraves, R., Carbone, L., Archidiacono, N., Albertson, D. G., Pinkel, D. *et al.* Large-scale variation among human and great ape genomes determined by array comparative genomic hybridization. *Genome Res.* **13,** 347–357 (2003).
6 Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305,** 525–528 (2004).
7 Sharp, A. J., Locke, D. P., McGrath, S. D., Cheng, Z., Bailey, J. A., Vallente, R. U. *et al.* Segmental duplications and copy-number variation in the human genome. *Am. J. Hum. Genet.* **77,** 78–88 (2005).

8 Marshall, C. R. & Scherer, S. W. Detection and characterization of copy number variation in autism spectrum disorder. *Methods Mol. Biol.* **838,** 115–135 (2012).
9 Swaminathan, G. J., Bragin, E., Chatzimichali, E. A., Corpas, M., Bevan, A. P., Wright, C. F. *et al.* DECIPHER: web-based, community resource for clinical interpretation of rare variants in developmental disorders. *Hum. Mol. Genet.* **21,** R37–R44 (2012).
10 Firth, H. V., Richards, S. M., Bevan, A. P., Clayton, S., Corpas, M., Rajan, D. *et al.* DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am. J. Hum. Genet.* **84,** 524–533 (2009).
11 Macdonald, J. R., Ziman, R., Yuen, R. K., Feuk, L. & Scherer, S. W. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res.* **42,** D986–D992 (2014).
12 Li, J., Yang, T., Wang, L., Yan, H., Zhang, Y., Guo, Y. *et al.* Whole genome distribution and ethnic differentiation of copy number variation in Caucasian and Asian populations. *PLoS ONE* **4,** e7958 (2009).
13 Lou, H., Li, S., Yang, Y., Kang, L., Zhang, X., Jin, W. *et al.* A map of copy number variations in Chinese populations. *PLoS ONE* **6,** e27341 (2011).
14 Zhang, Y. B., Li, X., Zhang, F., Wang, D. M. & Yu, J. A preliminary study of copy number variation in Tibetans. *PLoS ONE* **7,** e41768 (2012).
15 Kanduri, C., Ukkola-Vuoti, L., Oikkonen, J., Buck, G., Blancher, C., Raijas, P. *et al.* The genome-wide landscape of copy number variations in the MUSGEN study provides evidence for a founder effect in the isolated Finnish population. *Eur. J. Hum. Genet.* **21,** 1411–1516 (2013).
16 Hanihara, K. Dual structure model for the population history of the Japanese. *Jpn Rev.* **2,** 1–33 (1991).
17 Japanese Archipelago Human Population Genetics C., Jinam, T., Nishida, N., Hirai, M., Kawamura, S., Oota, H. *et al.* The history of human populations in the Japanese Archipelago inferred from genome-wide SNP data with a special reference to the Ainu and the Ryukyuan populations. *J. Hum. Genet.* **57,** 787–795 (2012).
18 Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S. F. *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* **17,** 1665–1674 (2007).
19 Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155,** 945–959 (2000).
20 Kumasaka, N., Yamaguchi-Kabata, Y., Takahashi, A., Kubo, M., Nakamura, Y. & Kamatani, N. Establishment of a standardized system to perform population structure analyses with limited sample size or with different sets of SNP genotypes. *J. Hum. Genet.* **55,** 525–533 (2010).
21 McCarroll, S. A., Kuruvilla, F. G., Korn, J. M., Cawley, S., Nemesh, J., Wysoker, A. *et al.* Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat. Genet.* **40,** 1166–1174 (2008).
22 Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y. *et al.* Origins and functional impact of copy number variation in the human genome. *Nature* **464,** 704–712 (2010).
23 Koike, A., Nishida, N., Yamashita, D. & Tokunaga, K. Comparative analysis of copy number variation detection methods and database construction. *BMC Genet.* **12,** 29 (2011).
24 Moon, S., Kim, Y. J., Hong, C. B., Kim, D. J., Lee, J. Y. & Kim, B. J. Data-driven approach to detect common copy-number variations and frequency profiles in a population-based Korean cohort. *Eur. J. Hum. Genet.* **19,** 1167–1172 (2011).
25 Vogler, C., Gschwind, L., Rothlisberger, B., Huber, A., Filges, I., Miny, P. *et al.* Microarray-based maps of copy-number variant regions in European and sub-Saharan populations. *PLoS ONE* **5,** e15246 (2010).
26 Shaikh, T. H., Gai, X., Perin, J. C., Glessner, J. T., Xie, H., Murphy, K. *et al.* High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. *Genome Res.* **19,** 1682–1690 (2009).
27 Fanciulli, M., Norsworthy, P. J., Petretto, E., Dong, R., Harper, L., Kamesh, L. *et al.* FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. *Nat. Genet.* **39,** 721–723 (2007).
28 Willcocks, L. C., Lyons, P. A., Clatworthy, M. R., Robinson, J. I., Yang, W., Newland, S. A. *et al.* Copy number of FCGR3B, which is associated with systemic lupus erythematosus, correlates with protein expression and immune complex uptake. *J. Exp. Med.* **205,** 1573–1582 (2008).
29 McKinney, C. & Merriman, T. R. Meta-analysis confirms a role for deletion in FCGR3B in autoimmune phenotypes. *Hum. Mol. Genet.* **21,** 2370–2376 (2012).
30 Yang, T. L., Chen, X. D., Guo, Y., Lei, S. F., Wang, J. T., Zhou, Q. *et al.* Genome-wide copy-number-variation study identified a susceptibility gene, UGT2B17, for osteoporosis. *Am. J. Hum. Genet.* **83,** 663–674 (2008).
31 Jakobsson, J., Ekström, L., Inotsume, N., Garle, M., Lorentzon, M., Ohlsson, C. *et al.* Large differences in testosterone excretion in Korean and Swedish men are strongly associated with a UDP-glucuronosyl transferase 2B17 polymorphism. *J. Clin. Endocrinol. Metab.* **91,** 687–693 (2006).
32 Pianezza, M. L., Sellers, E. M. & Tyndale, R. F. Nicotine metabolism defect reduces smoking. *Nature* **393,** 750 (1998).
33 Miyamoto, M., Umetsu, Y., Dosaka-Akita, H., Sawamura, Y., Yokota, J., Kunitoh, H. *et al.* CYP2A6 gene deletion reduces susceptibility to lung cancer. *Biochem. Biophys. Res. Commun.* **261,** 658–660 (1999).
34 Lv, J., Yang, Y., Zhou, X., Yu, L., Li, R., Hou, P. *et al.* FCGR3B copy number variation is not associated with lupus nephritis in a Chinese population. *Lupus* **19,** 158–161 (2010).
35 Xue, Y., Sun, D., Daly, A., Yang, F., Zhou, X., Zhao, M. *et al.* Adaptive evolution of UGT2B17 copy-number variation. *Am. J. Hum. Genet.* **83,** 337–346 (2008).
36 Oscarsona, M., McLellana, R. A., Gullsteènb, H., Yuec, Q. Y., Langd, M. A., Bernale, M. L. *et al.* Characterisation and PCR-based detection of a CYP2A6 gene deletion found at a high frequency in a Chinese population. *FEBS Lett.* **448,** 105–110 (1999).
37 Hammer, M. F. & Horai, S. Y chromosomal DNA variation and the peopling of Japan. *Am. J. Hum. Genet.* **56,** 951–962 (1995).
38 Shinka, T., Tomita, K., Toda, T., Kotliarova, S. E., Lee, J., Kuroki, Y. *et al.* Genetic variations on the Y chromosome in the Japanese population and implications for modern human Y chromosome lineage. *J. Hum. Genet.* **44,** 240–245 (1999).
39 Dulik, M. C., Zhadanov, S. I., Osipova, L. P., Askapuli, A., Gau, L., Gokcumen, O. *et al.* Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native Americans and Indigenous Altaians. *Am. J. Hum. Genet.* **90,** 229–246 (2012).
40 Oppenheimer, S. Out-of-Africa, the peopling of continents and islands: tracing uniparental gene trees across the map. *Philos. Trans. R. Soc. Lond. Ser. B* **367,** 770–784 (2012).
41 van den Berg, M., van Maarle, M., van Wely, M. & Goddijn, M. Genetics of early miscarriage. *Biochim. Biophys. Acta* **1822,** 1951–1959 (2012).
42 Rajcan-Separovic, E., Diego-Alvarez, D., Robinson, W. P., Tyson, C., Qiao, Y., Harvard, C. *et al.* Identification of copy number variants in miscarriages from couples with idiopathic recurrent pregnancy loss. *Hum. Reprod.* **25,** 2913–2922 (2010).
43 Bohnekamp, J. & Schoneberg, T. Cell adhesion receptor GPR133 couples to Gs protein. *J. Biol. Chem.* **286,** 41912–41916 (2011).
44 Perry, G. H., Dominy, N. J., Claw, K. G., Lee, A. S., Fiegler, H., Redon, R. *et al.* Diet and the evolution of human amylase gene copy number variation. *Nat. Genet.* **39,** 1256–1260 (2007).
45 1000 Genomes Project Consortium, Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M. *et al.* An integrated map of genetic variation from 1092 human genomes. *Nature* **491,** 56–65 (2012).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)

SCIENTIFIC REPORTS



OPEN

SUBJECT AREAS:
GENETIC ENGINEERING
CARTILAGE

Received
21 February 2014

Accepted
3 June 2014

Published
23 June 2014

Correspondence and requests for materials should be addressed to M.I. (inui-m@ncchd.go.jp) or S.T. (takada-s@ncchd.go.jp)

Rapid generation of mouse models with defined point mutations by the CRISPR/Cas9 system

Masafumi Inui[1], Mami Miyado[2], Maki Igarashi[2], Moe Tamano[1], Atsushi Kubo[3], Satoshi Yamashita[1], Hiroshi Asahara[1,3,4,5], Maki Fukami[2] & Shuji Takada[1]

[1]Department of Systems BioMedicine, National Research Institute for Child Health and Development, Tokyo 157-8535, Japan, [2]Department of Molecular Endocrinology, National Research Institute for Child Health and Development, Tokyo 157-8535, Japan, [3]Department of Systems BioMedicine, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo 113-8510, Japan, [4]CREST, Japan Science and Technology Agency (JST), Saitama 332-0011, Japan, [5]Department of Molecular and Experimental Medicine, The Scripps Research Institute, 10550 N. Torrey Pines Rd., La Jolla, CA 92037, USA.

Introducing a point mutation is a fundamental method used to demonstrate the roles of particular nucleotides or amino acids in the genetic elements or proteins, and is widely used in *in vitro* experiments based on cultured cells and exogenously provided DNA. However, the *in vivo* application of this approach by modifying genomic loci is uncommon, partly due to its technical and temporal demands. This leaves many *in vitro* findings un-validated under *in vivo* conditions. We herein applied the CRISPR/Cas9 system to generate mice with point mutations in their genomes, which led to single amino acid substitutions in proteins of interest. By microinjecting gRNA, hCas9 mRNA and single-stranded donor oligonucleotides (ssODN) into mouse zygotes, we introduced defined genomic modifications in their genome with a low cost and in a short time. Both single gRNA/WT hCas9 and double nicking set-ups were effective. We also found that the distance between the modification site and gRNA target site was a significant parameter affecting the efficiency of the substitution. We believe that this is a powerful technique that can be used to examine the relevance of *in vitro* findings, as well as the mutations found in patients with genetic disorders, in an *in vivo* system.

Introducing point mutations is a widely used experimental approach to evaluate the roles of specific nucleotides or amino acids in the function of the genetic elements or in proteins of interest. Using this technique, various nucleotides and amino acids have been proved to be indispensable for promoter or enhancer activity, as well as for the function or regulation of enzymes, transcription factors and signaling molecules. However, these findings were achieved mainly by *in vitro* experiments using exogenously provided DNA, such as plasmids or virus vectors, and the *in vivo* introduction of defined point mutations at endogenous genomic loci has been uncommon, largely because of the technical and temporal costs associated with such techniques. Thus, the *in vivo* context or relevance of *in vitro* findings has remained unexplored in many cases.

The CRISPR/Cas9 (clustered regularly interspaced short palindromic repeat/CRISPR-associated 9) system is a recently developed genome-engineering tool based on the bacterial CRISPR immune system, in which guide RNA (gRNA) recruits the Cas9 nuclease to the target locus in the genome by sequence complementarity, and induces double strand breaks (DSBs)[1,2]. These DSBs cause small insertions or deletions (indels) following non-homologous end-joining (NHEJ) repair, or can be utilized to introduce defined sequence modifications through a homology-dependent repair (HDR) mechanism[1,2]. While the use of HDR-dependent genomic engineering has been reported in *in vitro* cell culture systems, whether this could be applied in *in vivo* systems to introduce defined point mutations and what parameters affect its efficiency were less explored.

In the present study, we applied the CRISPR/Cas9 system to create a single amino acid substituted mouse model. By microinjecting synthesized RNAs and single-strand oligodeoxynucleotide (ssODN) donor sequences into mouse zygotes, we were able to introduce defined point mutations in the mouse genome, which led to single amino acid substitutions in the proteins of interest, within as short as one to two months. Using this technique, we can now evaluate the *in vivo* relevance of particular nucleotides/amino acids with a low cost and within a short time, which facilitates the elucidation of the *in vivo* context of the findings in cultured cell systems, as well as the *in vivo* screening for the relevant mutation among those found in human patients with various diseases.

Results

Substitution by single gRNA and wild-type hCas9. To generate an amino acid substituted mouse model using the CRISPR system, we first took the simplest approach: microinjecting single gRNA, wild type (WT) hCas9 mRNA and a ssODN with a single base mismatch to the genomic sequence, which caused single amino acid conversion, into mouse zygotes. As a target, a c.274C>T mutation of the *Steroidogenic Factor 1* gene (*Sf-1*, also known as *Ad4BP* or *Nr5a1*)[3], which causes a p.R92W amino acid substitution in the SF-1 protein, was chosen, and gRNAs and ssODNs were designed as shown in Figure 1A.

To examine whether the sequence or location of the gRNAs could affect the efficiency of the substitution, we designed three independent gRNAs (Sf-1gRNA1, 2 and 3), which had their Protospacer-Adjacent Motif (PAM) sequence 52 or 7 bases upstream, or 12 bases downstream, of the target point mutation (Fig. 1A). Three independent ssODNs were also designed, placing the mismatched nucleotide or PAM sequence of gRNA1 and gRNA2 in their center, respectively (Fig. 1A).

Various combinations of gRNAs and ssODNs (Table 1) were mixed with hCas9 mRNA and microinjected into pronuclear stage embryos. The injected zygotes were transferred to pseudo-pregnant females at the two-cell stage, and the obtained pups were examined for the sequence around the *Sf-1* locus. As shown in Table 1, we found total 26 mice with the C>T substitution generated from three out of four experiments (experiments 2, 3 and 4), and 7 of them had biallelic substitutions (Fig. 1B and Table 1). These results show that microinjecting gRNA, WT hCas9 mRNA and the ssODN donor into mouse zygotes is sufficient to introduce a defined single base substitution in their genome.

Of note, while the frequencies of DSB caused by hCas9/gRNAs, which are indicated by the frequencies of mutated alleles, are comparable among 4 experiments (68–89%, $P = 0.75$) (Table 1), the frequencies of designated substitutions vary significantly (0–29.6%, $P < 0.01$) (Table 1). This lack of a correlation between the frequency of the mutation and the substitution to occur suggests that there may be an optimal rate of DSB occurance to introduce HDR-dependent sequence incorporation (such like too frequent DSB might be unfavorable for HDR), or that the distance and/or the mutual position between the substitution target site and gRNA PAM sequences (surrogate DSB point) could affect the substitution efficiency. Our subsequent experiments support the latter possibility (see below). To explore the effect of mutual position between substitution target site and DSB point on substitution efficiency, we designed two additional ssODNs which have similar sequence to ssODN1, but carry point mutations in 36 base upstream or 12 base downstream to the PAM sequence of Sf-1 gRNA3 (Supplementary Fig. S1). ssODNs were microinjected into mouse zygotes with hCas9 mRNA and Sf-1 gRNA3, and the frequency of the designated substitution and the mutated allele in those mice were compared to experiment 4 (Supplementary Table S1). As a result, while 80–95% of mice had mutated alleles in all three experiments, the frequency of the substituted alleles differs substantially: 29.6% with ssODN3, 20.0% with ssODN4, and 0% with ssODN5 (Supplementary Table S1). These results indicate that the probability of the mutation in donor DNA to be incorporated into the genome declines with the distance between the DSB point and the mutation. In addition, the mutation which is at the downstream (3′ side) of PAM sequence showed slightly lower frequency of substitution than the mutation with the same distance from the PAM sequence but is at the upstream (5′ side). Together, these results show that the mutual location of gRNA target site and substitution target site is an important parameter for the efficient introduction of the designated point mutation.

Although there was an efficient generation of targeted point mutations, we observed that many mice with the C>T substitution also received additional random mutations in the flanking regions (Fig. 1C and Table 1). Moreover, it has previously been reported that single gRNA/WT hCas9 may introduce undesired mutations at the genomic loci with similar sequences to the target site (off-target effect)[4–6], which prompted us to test the double nicking strategy[5] to induce more accurate genome modifications in mouse zygotes.

Substitution using two gRNAs and hCas9 D10A (double nicking). Combining a pair of gRNAs with mutated hCas9 (hCas9 D10A), double nicking cause DSB only at the genomic loci where both of the gRNAs bind at an appropriate distance, thus making it highly specific to the target site[5]. To examine whether this set-up could be used for the HDR-dependent genome editing *in vivo*, we aimed to induce the c.1187A>G mutation of the *Sox9* gene, which results in a p.K396R amino acid substitution in the SOX9 protein. Lysine 396 is an evolutionary-conserved post-translational modification site of the SOX9 protein, but its *in vivo* relevance has not been fully examined[7,8]. We designed two gRNAs with a five-base offset where their PAM sequences are positioned about 40 bases upstream and three bases downstream of the A>G substitution site (Fig. 2A). Two ssODNs were designed with several silent mutations within the recognition sequence of gRNA2 to avoid repetitive digestions. An additional silent mutation to eliminate the *Pvu*II site was also designed to facilitate the genotyping (Fig. 2A). Sox9-ssODN1 causes both silent mutations and c.1187A>G substitution (K-to-R substitution), while Sox9-ssODN2 causes only silent mutations (K-to-K substitution) (Fig. 2A, lower panel). The two gRNAs, hCas9 D10A mRNA and 2 ssODNs were mixed and microinjected into zygotes to generate amino acid substituted mice (by Sox9-ssODN1) and negative control mice (by Sox9-ssODN2) simultaneously.

The obtained pups were genotyped as before, and we found that five pups out of 24 (20.8%) with the designated substitution, two of which had a K-to-R substitution (Fig. 2B) and three of which had a K-to-K substitution (Table 2). A frame shift mutation caused by an indel was observed in only one mouse. Thus, double nicking, together with donor ssODN, can also introduce defined point mutations in mouse zygotes. The frequency of the substitution was lower than that of the single gRNA/WT hCas9 condition shown in previous section, but was sufficient to generate mice with two genotypes in a single microinjection experiment by mixing different donor DNAs. Of note, a *Pvu*II site substitution was never observed in the mice with upstream substitutions (Fig. 2B and data not shown). This, together with the results of the experiment 4–6 with *Sf-1*, suggest that only a part of the sequence of the donor DNA near the DSB site will be incorporated through the CRISPR-mediated HDR pathway. Therefore, it is important to place the gRNA target site close enough to the substitution site to achieve efficient substitution.

SOX9-K396R gRNA2 and WT hCas9 mRNA with ssODNs were also microinjected into zygotes to compare the efficiency with double nicking (Table 2). As shown in Table 2, single gRNA with WT hCas9 resulted in a higher mutation frequency (76.9%) as well as substitution frequency (34.6%), and as in the case of *Sf-1*, biallelic substitution and additional random mutations were also observed (Supplementary Fig. S2 and Table 2).

Finally, we compared the off-target effects under the two conditions. Five putative off-target sites of SOX9-K396R gRNA2 were computationally predicted[6], and the sequences of those sites in gRNA/hCas9-injected mice were analyzed (Fig. 3A). As a result, we did not detect any mutations in the total of 250 loci (five off-target sites in 50 mice) examined (Fig. 3B and Table 3), indicating that both single gRNA with WT hCas9 and double nicking introduced designated genetic modification in highly specific manner. To examine if this high specificity could also be observed with other gRNAs, off-target effect of Sf-1 gRNAs were also analyzed. We focused on the mice which received designated substitution in first 4 experiments (total 26 mice shown in Table 1) and examined total of 121 loci, and did not find any off-target mutation (Supplementary Fig. S3 and

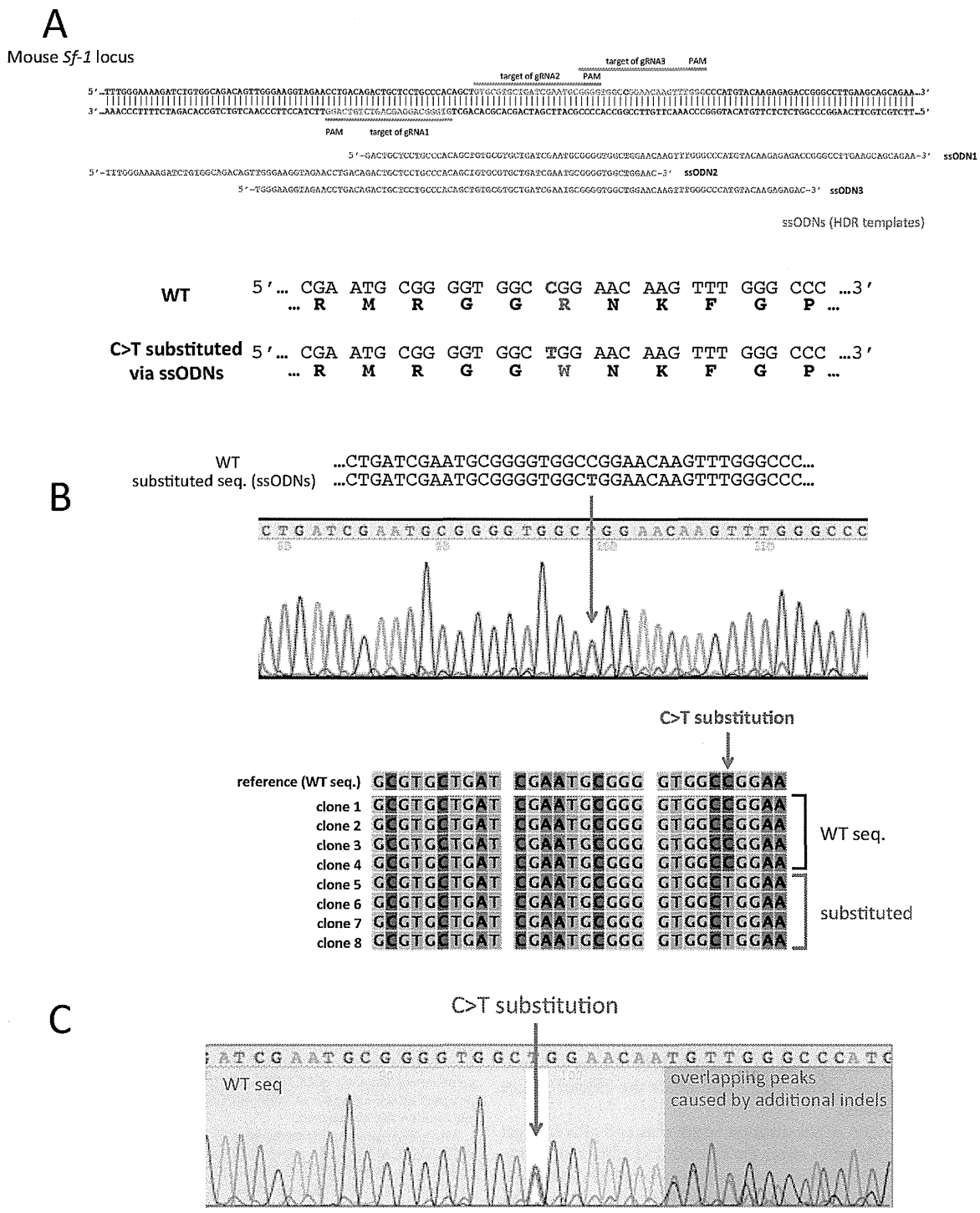


Figure 1 | Introduction of the C>T substitution in *Sf-1* locus. (A) A schematic illustration showing the locations of the gRNAs and ssODNs, along with the mouse *Sf-1* locus. Blue bars and letters indicate the position of the gRNA targets with red bars highlight the PAM sequences. ssODNs are shown in green letters at the corresponding position to the *Sf-1* locus. Red letters indicate the substitution target site in the *Sf-1* locus and corresponding mismatched nucleotides in the ssODNs. The deduced amino acid sequences from wild type (WT) and substituted sequences are shown at the bottom, with red letters indicating the target and the results of the substitution. (B) The results of the sequence analysis of the *Sf-1* locus of a representative mouse, with a monoallelic C>T substitution. The *Sf-1* locus around the gRNA target site was PCR amplified and directly sequenced (Upper panel), or cloned into the plasmid and sequenced independently (Lower panel). The upper panel shows the electropherogram of the direct sequencing results, in which one allele received a C>T substitution. WT and donor DNA sequences are shown on the top, with red letters indicating the mutated nucleotide. The red arrow indicates the overlapping peak caused by the monoallelic substitution. The lower panel shows the sequence alignment of eight independent clones. The eight sequences are separated into two types: the WT sequence and substituted sequences. (C) The results of the direct sequencing of the *Sf-1* locus around the gRNA target site of another representative mouse obtained in experiment 4 shown in Table 1, in which one alleles received C>T substitutions, followed by additional mutation(s) that caused continuous overlapping peaks.

Table 1 | Summary of experiments targeting the *Sf-1* locus

	Microinjection		Survival of zygotes			Result of genotyping				Ratio of genotyping	
Experiment	gRNA	ssODN	2 cell embryo/ injected zygotes	Transferred	Genotyped	Monoallelic substitution (+indel)	Biallelic substitution (+indel)	Indel	WT	Mouse with substituted alleles (%)	Mouse with mutated alleles (%)
1	1	1	128/256	128	39	0 (0)	0 (0)	31	8	0 (0)	31 (79.5)
2	1	2	101/203	101	37	1 (1)	0 (0)	32	4	1 (2.7)	33 (89.2)
3	2	3	124/169	124	51	8 (7)	1 (1)	26	16	9 (17.6)	35 (68.6)
4	3	1	157/217	144	54	10 (7)	6 (5)	28	10	16 (29.6)	44 (81.5)

Supplementary Table S2). Together, these results suggest the ratio of on/off-target mutation is substantially high in this experimental set up.

Discussion

Introducing point mutations is widely used technique in cultured cell based experiments, but uncommon in *in vivo* system partly because of its temporal and technical demands. The standard approach to examine the function of particular nucleotides or amino acids in *in vivo* background has been to create a knock-in or transgenic mouse that carries cDNA or other genetic element with defined point mutation(s). However these techniques require long and laborious procedures such as constructing targeting vectors, cloning of ES cell and/or establishing multiple transgenic mouse lines. Our method overcome those difficulties and possible problems and is ideal to observe the effect of subtle genetic modification under physiological condition.

The CRISPR/Cas9 system has been used to introduce defined genetic modification in cultured cell system[5,9], but whether this could be applied to *in vivo* system and what parameter(s) is influential for its efficiency was less explored. We demonstrated that the CRISPR/Cas9 system, both single gRNA/WT hCas9 and double nicking set-ups, was effective in mouse zygotes with the former having slightly higher efficiency. In our best condition, microinjection of single gRNA with WT hCas9 introduced the point mutation in more than 30% of mice examined, in which 10% of them received biallelic modification, allowing researcher to analyze the mouse model in the founder generation (Note that the possible mosaicism should be considered in analyzing the phenotype of founder generation and we believe that the confirmation in descendant generation is indispensable to conclude genotype-phenotype relationship). We observed higher efficiency with single gRNA with WT hCas9 than double nicking set-up, which is expected because the latter set-up requires two gRNA-hCas9 complexes to be recruited simultaneously to the target site to introduce DSB while one complex is enough for the former. In addition, to inject the same amount of RNAs in total, the concentration of each gRNA and hCas9 mRNA was lower in our double nicking set-up, which might also affect the efficiency (see Methods for the details).

The mutual position of substitution target sites and gRNA target sites (DSB points) seems to affect the substitution efficiency. Using same gRNA and ssODNs with various point mutations, we found that the mutations which are closer to the gRNA target site is incorporated more often than that are farther from DSB point (Supplementary Fig. S1 and Supplementary Table S1). This is in line with the previous report in which the efficiency of gene conversion was examined using plasmid donor DNA and restriction enzyme in cell culture system, and found that the frequency of the substitution declines with the distance from the digestion site[10]. Whether the frequency of substitution varies by using sense (i.e. the same strand of gRNA target) or anti-sense strand sequence for ssODN could be another point to be examined in the future studies.

Surprisingly, we did not observe any off-target effect in our experiments. The total of 371 loci were examined but no indels were detected. This was an unexpected result, since frequent off-target mutagenesis has been reported in human cells[4,5], but is consistent with a rare off-target effect observed in mouse zygotes microinjected with hCas9 and gRNA plasmids[11]. These results do not exclude off-target effects in general, but imply that the cultured cells and mouse zygotes have distinct risks for off-target effects, possibly due to the different stability of the CRISPR/Cas9 components and/or the nuclear/chromatin dynamics between stable cell lines and zygotes under early developmental process. Further accumulation of data is necessary to understand the kinetics of CRISPR mediated genome modification to optimize the experimental parameters for obtaining highest efficiency with minimum off-target effect in various back ground.

In summary, we have established an experimental system to introduce defined point mutations into the mouse genome using the CRISPR/Cas9 system, which enabled us to generate an amino acid substituted mouse model at a low cost and in a short time. The WT hCas9 and double nicking strategy were both effective and the distance between the DSB site and substitution site seems to be an important parameter for determining the substitution efficiency. We believe this technique will be useful to reveal the *in vivo* relevance of particular nucleotides/amino acids identified in *in vitro* systems, as well as to reproduce the genetic variations found in patients of genetic disorders in mouse models.

Methods

Plasmids. hCas9, hCas9 D10A and the gRNA cloning vector[1] were purchased from Addgene (Plasmid ID #41815, #41816 and #41824, respectively). Since the original gRNA cloning vector lacks the partial sequence of the U6 promoter and gRNA scaffold, we first modified it by adding the following sequence: 5′-GTGGAA-AGGACGAAACACCGGCTAGCAGGCCTATCGATGTTTTAGAGCTAGAA-ATAGC-3′ into the *Afl*II site to fill the missing sequence and to facilitate further cloning. gRNA expression vectors were constructed by inverse PCR using this modified vector and the primer pairs shown in the Table below. The PCR products were *Dpn*I digested and used for the transformation of *E. coli* (DH5a). The sequences of the obtained constructs were validated by sequencing.

gRNA design. The gRNAs were designed by searching for "GG" or "CC" sequences near the point mutation target sites, and were defined as N(21)GG or the reverse complement sequence of CCN(21). The gRNA pairs used for double-nicking were designed by searching for CCN(40–48)GG sequences, and the pair closest to the substitution site was chosen.

The gRNAs used in this study were: Sf-1 gRNA1 5′-GTGGGCAGGAGC-AGTCTGTCAGG-3′, Sf-1 gRNA2 5′-GTGCGTGCTGATCGAATGCGGGG-3′, Sf-1 gRNA3 5′-GGGGTGGCCGGAACAAGTTTGGG-3′, SOX9-K396R gRNA1 5′-GCCTGGCTCGCTGCTCAGCGTGG-3′, SOX9-K396R gRNA2 5′-CCAGCGAAC-GCACATCAAGACGG-3′.

RNA synthesis. Template DNAs for *in vitro* transcription were generated by PCR amplification of the ORFs of hCas9 and hCas9 D10A, as well as gRNAs (i.e. guide sequences and scaffold sequences) from each plasmid using PrimeSTAR (TaKaRa) and the primer sequences shown below (the T7 RNA polymerase recognition sequence was attached to the 5′ end of the fw primers). The PCR products were purified and used for the *in vitro* RNA synthesis using mMessage mMachine T7 kit (lifetechnologies). The reaction scale was doubled, and the reaction time was extended to over-night to obtain a sufficient amount of gRNA. The synthesized RNAs were purified using an RNeasy mini kit (Qiagen) with an additional ethanol precipitation.

ssODN. The synthesized single stranded oligonucleotides (110 bases, PAGE purified) were purchased from fasmac.

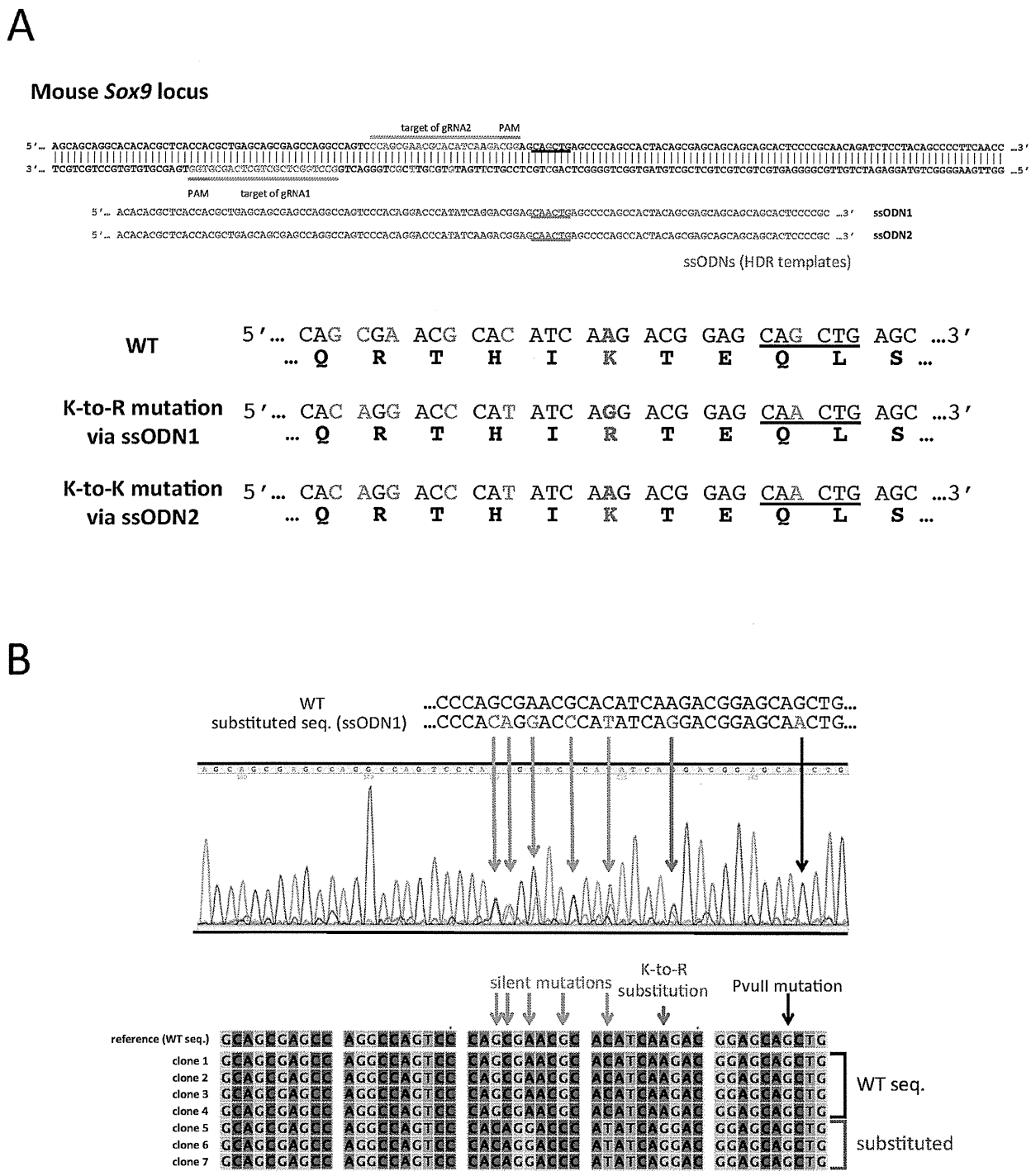


Figure 2 | Introduction of the K396R substitution in *Sox9* locus. (A) A schematic illustration demonstrating the locations of the gRNAs and ssODNs along with the mouse *Sox9* locus. The blue bars and letters indicate the positions of gRNA targets, while red bars highlight the PAM sequences. The sequences of ssODNs are shown in green letters at the corresponding positions to the *Sox9* locus. Red letters indicate the substitution target site in the *Sox9* locus and the corresponding mismatched nucleotides in the ssODNs. The deduced amino acid sequences from wild type (WT) and substituted sequences are shown at the bottom, with red letters indicating the target and the results of the substitution. Orange letters in the sequences are the targets of silent mutations, which do not cause amino acid conversion. The *Pvu*II site is underlined in the *Sox9* loci which had mutated in ssODNs. (B) The results of the sequence analysis of the *Sox9* locus of a representative mouse, which had a monoallelic K-to-R substitution. The *Sox9* locus around the gRNA target site was PCR amplified and directly sequenced (Upper panel), or was cloned into the plasmid and sequenced independently (Lower panel). The upper panel shows the electropherogram of the direct sequencing result of a mouse obtained in experiment 1 shown in Table 2, in which one allele received a K-to-R substitution. The WT and donor DNA sequences are shown on the top, with orange and red letters indicating mutated nucleotides. The orange and red arrows indicate the overlapping peaks caused by monoallelic substitutions, and a black arrow indicates the non-overlapping peak despite the fact that the ssODN (donor DNA) had a mutated nucleotide. The lower panel shows the sequence alignment of seven independent clones. The seven sequences were separated into two types: the WT and substituted sequences.

Table 2 | Summary of experiments targeting the *Sox9* locus

	Microinjection			Survival of zygotes			Result of genotyping				Ratio of genotype	
Experiment	gRNA	hCas9	ssODN	2 cell embryo/ injected zygotes	Transferred	Genotyped	Monoallelic substitution (+indel)	Biallelic substitution (+indel)	Indel	WT	Mouse with substituted alleles (%)	Mouse with mutated alleles (%)
1	1 & 2	D10A	1 & 2	144/188	144	24	5 (0)	0 (0)	1	18	5 (20.8)	6 (25.0)
2	2	WT	1 & 2	116/139	90	26	6 (1)	3 (1)	11	6	9 (34.6)	20 (76.9)

Sf-1 ssODN1 5′-GACTGCTCCTGCCCACAGCTGTGCGTGCTGATCGAATGCGGGGTGGCTGGAACAAGTTTGGGCCCATGTACAAGAGAGACCGGGCCTTGAAGCAGCAGAA-3′, Sf-1 ssODN2 5′-TTTGGGAAAAGATCTGTGGCAGACAGTTGGGAAGGTAGAACCTGACAGACTGCTCCTGCCCACAGCTGTGCGTGCTGATCGAATGCGGGGTGGCTGGAAC-3′, Sf-1 ssODN3 5′-TGGGAAGGTAGAACCTGACAGACTGCTCCTGCCCACAGCTGTGCGTGCTGATCGAATGCGGGGTGGCTGGAACAAGTTTGGGCCCATGTACAAGAGAGAC-3′, Sf-1 ssODN4 5′-GACTGCTCCTGCCCACAGCTGTGCGTGCTGATCGAATGCGGGGTGGCCGGAACAAGTTTGGGCCCATGTATAAGAGAGACCGGGCCTTGAAGCAGCAGAA-3′, Sf-1 ssODN5 5′-GACTGCTCCTGCCCACAGCTGTGTGCTGATCGAATGCGGGGTGGCCGGAACAAGTTTGGGCCCATGTACAAGAGAGACCGGGCCTTGAAGCAGCAGAA-3′, Sox9-K396R ssODN1 5′-ACACACGCTCACCACGCTGAGCAGCGAGCCAGGCCAGTCCCACAGGACCCATATCAGGACGGAGCAACTGAGCCCCAGCCACTACAGCGAGCAGCAGCAGCACTCCCCGC-3′, Sox9-K396R ssODN2 5′-ACACACGCTCACCACGCTGAGCAGCGAGCCAGGCCAGTCCCACAGGACCCATATCAAGACGG-

A

	chromosome	position	strand	sequence	# mismatches
Sox9 locus	chr11	112,785,156-112,785,178	+	CCAGCGAACGCACATCAAGACGG	0
OT-1	chrX	78,583,896-78,583,918	+	CCAGAGAACCCACATCAAGAGGG	2
OT-2	chr8	123,777,951-123,777,973	+	CCAGCGAGGGCACACCAAGAAGG	3
OT-3	chr6	15,067,966-15,067,988	+	CCAGCTAAAGCACATCTAGAGAG	3
OT-4	chr15	97,443,330-97,443,352	-	CCAGCGAGCTCACTTCAAGACAG	3
OT-5	chr13	72,438,524-72,438,546	+	CCAGTGAAAGCACATGAAGATGG	3

B

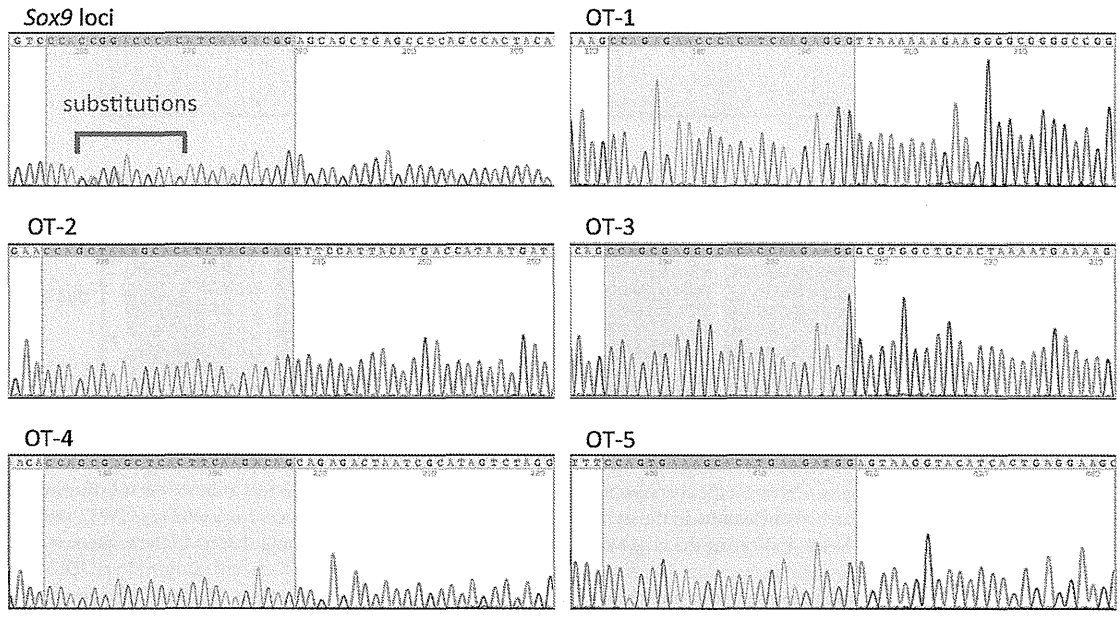


Figure 3 | Off-target analysis of SOX9-K396R gRNAs. (A) The sequences and positions of possible off-target sites for SOX9-K396R gRNA2. OT: Off-target site. (B) The results of the sequence analysis of the target site and off-target sites. Direct sequencing of the PCR-amplified *Sox9* locus and five putative off-target sites from single representative mouse are shown as representative results. Only the sequence of the *Sox9* locus showed overlapping peaks caused by substitution. The gRNA target and possible off-target sequences were shown at the fourth to twenty-sixth bases, as indicated by the darker background color.

Table 3 | Summary of off-target effects by SOX9-K396R gRNAs

Target sites	Mutation rate with WT Cas9 (%)	Mutation rate with double nicking (%)
Sox9 locus	20/26 (76.9)	6/24 (25.0)
OT-1	0/26 (0)	0/24 (0)
OT-2	0/26 (0)	0/24 (0)
OT-3	0/26 (0)	0/24 (0)
OT-4	0/26 (0)	0/24 (0)
OT-5	0/26 (0)	0/24 (0)

OT: Off-target site.
Mutation rate: No. of mice with a mutated allele (including indels and substitutions)/No. of mice genotyped.

AGCAACTGAGCCCCAGCCACTACAGCGAGCAGCAGCAGCACTCCC-CGC-3′.

Microinjection. The microinjection of mouse zygotes was performed as described before[12,13]. Essentially, mouse zygotes were obtained by mating superovulated BDF1 females and WT BDF1 males (Sankyo lab service). RNAs and ssODNs were mixed just before microinjection into the cytoplasm or pro-nuclei of zygotes, and the injected embryos were incubated at 37°C until they were transferred into pseudo-pregnant females at the two-cell stage. The concentration of injected RNAs was always kept at 500 ng/μl in total. For the single gRNA/WT Cas9 condition, gRNA and hCas9 mRNA were mixed at a 1 : 1 ratio, and thus the final concentration was 250 ng/μl each, and for the double nicking condition, the gRNAs and hCas9 mRNA were mixed at 1 : 1 : 1 ratio, and thus a final concentration of 167 ng/μl each. The concentration of injected ssODNs was final 100 ng/μl. The protocols for animal experiments were approved by the Animal Care and Use Committee of the National Research Institute for Child Health and Development (Permit Numbers: A2004-003-C09, A2009-002-C04).

Genotyping. Genomic DNA was extracted from the tail tips of pups, and the genomic sequences around the gRNA target sites were PCR amplified using the primers shown below. The obtained PCR products were treated by ExoSAP-IT (USB) and sequenced directly, or were cloned into the plasmid and sequenced.

The criteria for the "mutation" and "substitution". The genomic sequences around the gRNA target sites were PCR amplified and served for direct sequencing (see above). The electropherograms of each sequence were classified into three categories: wild type (no overlapping peaks), mutated (continuous overlapped peaks, which started from the gRNA target site) and substituted (overlapped peak(s) at the designated position(s)) (Supplementary Fig. S2). The sequences were also aligned with the wild type sequence to check whether there was a homozygous mouse with identical indels on both alleles (which also produce a no overlapping peak pattern). PCR products were cloned into the plasmid, and multiple clones were sequenced to determine the sequences of each allele when necessary.

Off-target analysis. Possible off-target sites for SOX9-K396R gRNA2 and Sf-1 gRNA 1, 2 and 3 were predicted by an online-based tool (http://www.genome-engineering.org/)[6]. Four or five putative off-target sites were found allowing a maximum of three to four bases of mismatches (Fig. 3A and Supplementary Fig. S3). The off-target sites were PCR amplified from the genomic DNA of gRNA-injected mice and analyzed by direct sequencing.

Primers. The primer sequences used in this study are shown below:

For vector construction: gRNA vector adopter fw 5′-GCTAGCAGGCCTATCGATGTTTTAGAGCTAGAAATAGCAAGTTAAAATA-AGGCTAGTCCGTTATC-3′, gRNA vector adopter rev 5′-ATCGATAGGCCTGCTAGCCGGTGTTTCGTCCTTTCCACAAGATATATAA-AGCCAAGAAATCGAAATAC-3′.

For gRNA cloning: SOX9-K396R gRNA1 fw 5′-CCTGGCTCGCTGCTCAGC-GGTTTTAGAGCTAGAAATAGCAAG-3′, SOX9-K396R gRNA1 rev 5′-AACCGCTGAGCAGCGAGCCAGGCGGTGTTTCGTCCTTTCCAC-3′, SOX9-K396R gRNA2 fw 5′-CAGCGAACGCACATCAAGAGTTTTAGAGCTAG-AAATAGCAAG-3′, SOX9-K396R gRNA2 rev 5′-AACTCTTGATGTGCGTTC-GCTGCGGTGTTTCGTCCTTTCCAC-3′, Sf-1-gRNA1 fw 5′-TGGGCAGGA-GCAGTCTGTCGTTTTAGAGCTAGAAATAGCAAG-3′, Sf-1-gRNA1 rev 5′-AACGACAGACTGCTCCTGCCCACGGTGTTTCGTCCTTTCCAC-3′, Sf-1-gRNA2 fw 5′-TGCGTGCTGATCGAATGCGGTTTTAGAGCTAGAAA-TAGCAAG-3′, Sf-1-gRNA2 rev 5′-AACCGCATTCGATCAGCACGCACGG-TGTTTCGTCCTTTCCAC-3′, Sf-1-gRNA3 fw 5′-GGGTGGCCGGAACAA-GTTTGTTTTAGAGCTAGAAATAGCAAG-3′, Sf-1-gRNA3 rev 5′-AACAAAC-TTGTTCCGGCCACCCCGGTGTTTCGTCCTTTCCAC-3′.

For RNA synthesis:

T7-Sf-1-gRNA1 5′-TTAATACGACTCACTATAGGTGGGCAGGAGCAGTCT-GTC-3′, T7-Sf-1-gRNA2 5′-TTAATACGACTCACTATAGGTGCGTGCTGA-TCGAATGCG-3′, T7-Sf-1-gRNA3 5′-TTAATACGACTCACTATAGGG-GGTGGCCGGAACAAGTTT-3′, T7-SOX9-K396R gRNA1 5′-TTAATACGA-CTCACTATAGGCCTGGCTCGCTGCTCAGCG-3′, T7-SOX9-K396R gRNA2 5′-TTAATACGACTCACTATAGGCAGCGAACGCACATCAAGA-3′, gRNA rev template for RNA synthesis 5′-AAAAGCACCGACTCGGTGCC-3′, T7-hCAS9 5′-TAATACGACTCACTATAGGGAGAATGGACAAGAAGTACTCCATTGG-3′, hCAS9-rev 5′-TCACACCTTCCTCTTCTTC-3′.

For genotyping: Sox9 exon 3 fw 5′-ACCAATACTTGCCACCCAAC-3′, Sox9 exon 3 rev 5′-CGGCTGCGTGACTGTAGTAG-3′, Sf-1 exon 4 fw 5′-TGGGGAATG-GTATAAGCGTGTG-3′, Sf-1 exon 4 rev 5′-CGTGCAGGCTAGGGGGTAAC-3′.

For the off-target analysis:

SOX9-K396R gRNA2 OT1 fw 5′-TACACACCCGAGTCCCTTTC-3′, SOX9-K396R gRNA2 OT1 rev 5′-ACCAAACAACACGGCCTTAG-3′, SOX9-K396R gRNA2 OT2 fw 5′-ATCTGACTTGGCGTGGAAAC-3′, SOX9-K396R gRNA2 OT2 rev 5′-GACCAGGACCTGTCGTCAAT-3′, SOX9-K396R gRNA2 OT3 fw 5′-GCC-AAGAGGAGGTAGCAGTG-3′, SOX9-K396R gRNA2 OT3 rev 5′-CACATCCC-CATAGGAAATGG-3′, SOX9-K396R gRNA2 OT4 fw 5′-GTCTCTTGGCCTC-TGCAATC-3′, SOX9-K396R gRNA2 OT4 rev 5′-AGCTCCACCCACAGAG-AGAA-3′, SOX9-K396R gRNA2 OT5 fw 5′-AGACAGAGCTGCTGCAAACA-3′, SOX9-K396R gRNA2 OT5 rev 5′-TGTCAACTACCGCACATGGT-3′, Sf-1 gRNA1 OT1 fw 5′-AGCAGAGAAGCAGGAGCAAG-3′, Sf-1 gRNA1 OT1 rev 5′-CCA-TCCCAAACTCAGCTGTT-3′, Sf-1 gRNA1 OT2 fw 5′-CTTGACTGCTTCTG-GGAAGG-3′, Sf-1 gRNA1 OT2 rev 5′-GCCCACCTGTGCCTTATCTA-3′, Sf-1 gRNA1 OT3 fw 5′-GACCATTGGAGGGCAAACTA-3′, Sf-1 gRNA1 OT3 rev 5′-GGCCAGGCTATAAACACCAA-3′, Sf-1 gRNA1 OT4 fw 5′-AGTGTGAGG-TCCCTGTTTGG-3′, Sf-1 gRNA1 OT4 rev 5′-AGGCTTAGGGATCTGGCATT-3′, Sf-1 gRNA1 OT5 fw 5′-CTTGCCTTTTCTCTGCCATC-3′, Sf-1 gRNA1 OT5 rev 5′-GCAGCCTGAAGGAAATGAAG-3′, Sf-1 gRNA2 OT1 fw 5′-TGGGAGAGT-TCCCTGATTTG-3′, Sf-1 gRNA2 OT1 rev 5′-GTCTCCAAACCCACCAGAAA-3′, Sf-1 gRNA2 OT2 fw 5′-ACTCCCTGGGACACTGTCTG-3′, Sf-1 gRNA2 OT2 rev 5′-TGCAAGAGTCCACACTACGG-3′, Sf-1 gRNA2 OT3 fw 5′-TTCCTAAGT-TGGCTCGCAGT-3′, Sf-1 gRNA2 OT3 rev 5′-CTTGGGTTTTTCATGGGCTA-3′, Sf-1 gRNA2 OT4 fw 5′-GATTAAAGGCGTGTGCCACT-3′, Sf-1 gRNA2 OT4 rev 5′-TCCCAGCCACATTCATGTTA-3′, Sf-1 gRNA3 OT1 fw 5′-GCACCGTCA-AGGAGAAAGAG-3′, Sf-1 gRNA3 OT1 rev 5′-GTTCCCCAGTCCTTCAACAA-3′, Sf-1 gRNA3 OT2 fw 5′-CACCAATTCAGCGACATCAG-3′, Sf-1 gRNA3 OT2 rev 5′-AAAGATGCTTGCACCTTGCT-3′, Sf-1 gRNA3 OT3 fw 5′-TGAGTCCAGC-CTCATCACAG-3′, Sf-1 gRNA3 OT3 rev 5′-CCCTGTCTCCAAACACCCTA-3′, Sf-1 gRNA3 OT4 fw 5′-GCTTGTCATGCTCAGTTCCA-3′, Sf-1 gRNA3 OT4 rev 5′-GACCTTTGGAAGAGCAGTCG-3′, Sf-1 gRNA3 OT5 fw 5′-TCCAGGAGACC-TGTTGCTCT-3′, Sf-1 gRNA3 OT5 rev 5′-AGGCTCCAAGACAGTGTGGT-3′.

Statistical analysis. Statistical analysis to assess the differences of the frequency of mutations and substitutions among experiments 1–4 (Table 1) was performed using the χ^2 test. We hypothesized that the expected frequency (calculated as total number of mice with mutated (or substituted) allele/total number of genotyped mice) is same among 4 experiments and esamined with χ^2 test.

1. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* **339**, 823–826 (2013).
2. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819–823 (2013).
3. Parker, K. L. & Schimmer, B. P. Steroidogenic factor 1: a key determinant of endocrine development and function. *Endocrine Reviews* **18**, 361–377 (1997).
4. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31**, 822–6 (2013).
5. Ran, F. A. *et al.* Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* **154**, 1380–9 (2013).
6. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
7. Akiyama, H. *et al.* The transcription factor Sox9 is degraded by the ubiquitin-proteasome system and stabilized by a mutation in a ubiquitin-target site. *Matrix Biol.* **23**, 5083–5093 (2005).
8. Lee, P. C. *et al.* SUMOylated SoxE factors recruit Grg4 and function as transcriptional repressors in the neural crest. *J. Cell Biol.* **198**, 799–813 (2012).
9. Wang, H. *et al.* One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell* **153**, 910–918 (2013).

10. Elliott, B., Richardson, C., Winderbaum, J., Nickoloff, J. A. & Jasin, M. Gene conversion tracts from double-strand break repair in mammalian cells. *Mol. Cell Biol.* **18**, 93–101 (1998).
11. Mashiko, D. *et al.* Generation of mutant mice by pronuclear injection of circular plasmid expressing Cas9 and single guided RNA. *Sci. Rep.* **3**, 3355 (2013).
12. Kato, T. *et al.* Production of Sry knockout mouse using TALEN via oocyte injection. *Sci. Rep.* **3**, 3136 (2013).
13. Takada, S. *et al.* Targeted Gene Deletion of miRNAs in Mice by TALEN System. *PLoS ONE* **8**, e76004 (2013).

Acknowledgments

This work was supported in part by the Grant from the National Center for Child Health and Development (Grant Number 25-1 for M. Inui and 24-3 for S.T.), Japan Society for the Promotion of Science KAKENHI (Grant Number 25871177 for M. Inui and Grant Number 25132713 for S.T.), the Ministry of Education, Culture, Sports, Science and Technology KAKENHI (Grant Number 23570265) to S.T.

Author contributions

M. Inui designed the substitution of *Sox9* and M. Igarashi. and M.F. defined the position of *Sf-1* substitution. A.K. constructed gRNA vectors. M.T. performed the microinjection. M. Inui, M.M. and S.Y. carried out the genotyping and the sequence analysis. M. Inui, H.A. and S.T. designed the project and wrote the manuscript.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Inui, M. *et al.* Rapid generation of mouse models with defined point mutations by the CRISPR/Cas9 system. *Sci. Rep.* 4, 5396; DOI:10.1038/srep05396 (2014).

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/