■ 質問 2 で「いる」に○をつけた方は、質問 3 へお進みください。

↓

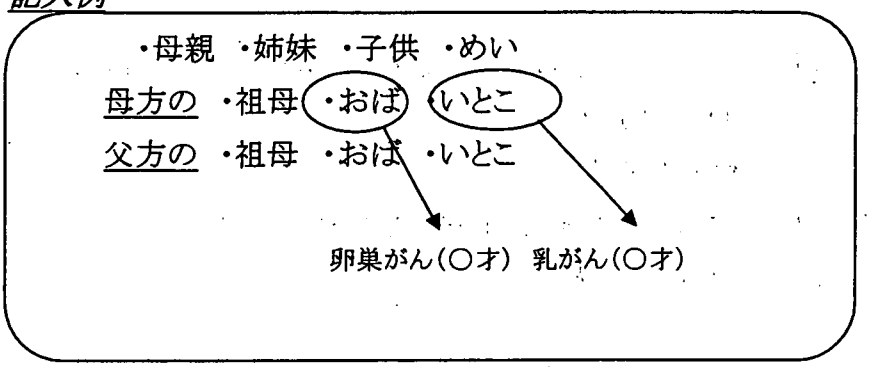
ご質問3. 乳がん、卵巣がんになられた方の、あなたとのご関係は？

※ 下記のように該当する方に○をつけ、乳がんか卵巣がんかを記入してください。

※ 複数いらっしゃる場合は、その人数だけ○をつけてください。

※ また乳がん、卵巣がんになられた時の年齢も書いてください。

記入例



・母親 ・姉妹 ・子供 ・めい
母方の ・祖母 ・おば ・いとこ
父方の ・祖母 ・おば ・いとこ

■ ご質問4. 質問 3 で、母親または母方のご家族に○をつけた方は以下の質問にお答えください。

母親のご兄弟・姉妹(あなたのおじ・おば)の人数を教えてください。

おじ(　　)人、おば(　　)人

母親のご兄弟・姉妹の子供(あなたのいとこ)の人数・性別を教えてください。

いとこ:男性(　　)人、女性(　　)人

母方のいとこの方が乳がんまたは卵巣がんであった場合、その方はあなたのおじ・おばどちらの子供ですか？

・おじの子供　・おばの子供

また、その母方のいとこの方のご兄弟・姉妹の数を教えてください。

兄弟(　　)人、姉妹(　　)人

■ ご質問 5. 質問 3 で、父方のご家族に○をつけた方は以下の質問にお答えください。

父親のご兄弟・姉妹(あなたのおじ・おば)の人数を教えてください。

おじ(　　)人、おば(　　)人

父親のご兄弟・姉妹の子供(あなたのいとこ)の人数・性別を教えてください。

いとこ:男性(　　)人、女性(　　)人

父方のいとこの方が乳がんまたは卵巣がんであった場合、その方はあなたのおじ・おばどちらの子供ですか?

・おじの子供　・おばの子供

また、その父方のいとこの方のご兄弟・姉妹の数を教えてください。

兄弟(　　)人、姉妹(　　)人

■ ご質問 6. 質問 3 でめいに○をつけた方は以下の質問にお答えください。

そのめいの方は、あなたのご兄弟の子ですかご姉妹の子ですか?

・兄弟の子　・姉妹の子

そのめいの方のご兄弟・姉妹の数を教えてください。

兄弟(　　)人、姉妹(　　)人

■ ご質問 7. まれに男性も乳がんにかかるといわれていますが、ご家族の男性で乳がんにかかった方はいらっしゃいますか?

・いる　・いない

→「いる」に○をつけた方、その男性とあなた様のご関係は?(○をおつけ下さい)

・ 父親 ・兄弟 ・祖父(母方・父方)・おじ(母方・父方)・おい(母方・父方)・いとこ(母方・父方)

以上で質問は終わりです。

ご協力まことにありがとうございました。

卵巣癌特異的染色体異常やヘテロ接合性消失のゲノム網羅的同定に関する研究

厚生労働省科学研究費補助金(第3次対がん総合戦略研究事業)

分担研究報告書

卵巣癌特異的染色体異常やヘテロ接合性消失のゲノム網羅的同定に関する研究

分担研究者　井ノ上 逸朗　東海大学医学部 基礎医学系　教授

研究要旨

卵巣癌や乳癌は、他の癌種と同様、細胞遺伝学的異常を伴うことが知られているものの、染色体不安定性に関わる遺伝要因が十分明らかになっているとはいえない。本年度は、SNP アレイとよばれるゲノム網羅性の高いマイクロアレイを用いて、卵巣癌ゲノム特異的染色体構造異常領域をゲノム全域から体系的に同定することを目的とした。卵巣癌–正常ゲノム対サンプルにおける比較解析から、卵巣癌ゲノムは、1コピー重複・欠失といったアレル特異的コピー数変化の出現頻度が非常に高いという特徴を有することが明らかになった。このような染色体不安定性に関する特徴は、BRAC1 変異陽性(ギャリアー)家族性卵巣癌、孤発性卵巣癌のいずれにおいても認められ、卵巣癌に共通する特性であると考えられる。来年度以降、乳癌においても、癌–正常対サンプル間の比較解析を行い、染色体不安定性パターンについての癌種による異同を調べるとともに、発癌感受性に関わる遺伝要因の同定を目指す。

A. 研究目的

卵巣癌や乳癌は、通常、染色体の異数化などの細胞遺伝学的異常を伴うことが知られている。その一方、癌ゲノムにおける従来の細胞遺伝学的研究からは、染色体不安定性に関わる遺伝要因が明らかにされているとはいえない。本研究では、SNP アレイとよばれるゲノム網羅性の高いマイクロアレイを用いて、癌ゲノム特異的染色体構造異常領域をゲノム全域から体系的に同定することを目的とする。特に、BRCA1 変異陽性(キャリアー)における家族性乳癌卵巣癌の発症機序解明を目指して、発症感受性と染色体構造異常との関連を調べる。本年度は、BRCA1 キャリアー家族性卵巣癌と孤発例との比較から、染色体不安定性という観点において卵巣癌ゲノムを特徴付けることを主たる目的とした。取得したデータは、感受性に関わる遺伝因子を探索するための基礎資料となる。

B. 研究方法

卵巣癌ゲノムにおけるアレル特異的変化も研究対象とするために、1塩基多型(SNP)とコピー数変化とを同時に分析することができる Genome-Wide Human SNP Array 5.0 (Affymetrix 社)を用いて解析を行った。卵巣癌および正常組織(43 対応対サンプル:孤発性 33 対、BRCA1 キャリアー家族性 10 対)からのゲノム DNA サンプルを用いて、ゲノム構造変化を高密度に分析した。また、卵巣癌罹患者の正常ゲノムにおける遺伝性のゲノム構造多型頻度を推定するために、BRCA1 キャリアー正常ゲノム 10 サンプルも解析に供した。SNP アレイデータに基づく SNP 遺伝子型の推定には、BRLMM-P アルゴリズム(Affymetrix 社)を用いた。アレイ上の SNP タイピング用および非多型プローブにおけるシグナル強度比較解析には、Partek Genomics Suite 6.3 ソフトウェア (Partek 社)を用いた。

(倫理面への配慮)

新潟大学および東海大学倫理審査委員会にて患者卵巣組織の採取、癌細胞および正常細胞由来 DNA を用いた研究について承認を得ている。個人情報管理者による適切な情報管理のもと研究を遂行しており、研究実施施設(東海大学医学部)には個人情報は一切存在しない。

C. 研究結果

卵巣癌ゲノム、正常ゲノムにおける SNP 遺伝子型コール率(SNP 数 440,794)は、それぞれ 98.36% (*N*=43)、99.34% (*N*=53)となり、信頼性の高いデータが取得されたと考えた。次いで、正常ゲノムにおけるゲノム構造多型の出現頻度につき、BRCA1 キャリアー群および孤発性群の間で比較したところ、BRCA1 キャリアー群で高頻度出現する傾向を認めたものの(38.9% vs. 24.2%)、統計学的な差異は見いだされなかった。

癌–正常対サンプルにおける解析から、非常に多くの卵巣癌ゲノム特異的構造変化を見いだした。興味深いことに、1コピー重複や欠失といったアレル特異的コピー数変化が、全体の約 88%を占めることが明らかになった。一方、BRCA1 キャリアー–孤発性の比較からは、癌ゲノムにおける構造変化パターンや変化数に顕著な差異を認めなかったものの、数多くの共通性が見いだされた。全 43 サンプルのうち、40%以上で共通した癌ゲノム特異的構造変化領域(0.1Mbp 以上)は全部で9カ所あり、その内訳は、1コピー重複5カ所(3q, 5p, 7q, 8q, 20p)、1コピー欠失4カ所(4q, 5q, 17p, 18q)であった。

D. 考察

卵巣癌ゲノムで観察された構造変化のうち、アレル特異的コピー数変化が大多数を占めるという知見は、SNP アレイを使用することで新規に見いだされたものである。従来の比較ゲノムハイブリダイゼーション(CGH)法ではアレル特異的変化を同定し得ないことを考え併せると、ゲノム構造変化解析への SNP アレイの有効性を示す知見であると考える。

また、このことは、卵巣癌ゲノム解析におけるアレル不均衡解析の重要性を示唆している。次年度は、特に uniparental disomy(片親ダイソミ

ー)といったコピー数変化を伴わないアレル不均衡状態を検出するための解析および実験的検証を進めるとともに、乳癌ゲノムに対しても同様の比較解析を行うことで、卵巣癌および乳癌ゲノムの不安定性について、アレル不均衡の立場から、遺伝学的に特徴付ける予定である。加えて、卵巣癌組織における遺伝子発現量解析結果とゲノム構造変化との組み合わせから、発癌に関わる遺伝子(遺伝領域)の同定を試みる。

E. 結論

卵巣癌ゲノムにおいては、アレル特異的コピー数変化(1コピー重複・欠失)の出現頻度が非常に高いことを明らかにした。

F. 健康危険情報

特記事項なし。

G. 研究発表

1)論文発表

1. Okada H, Tajima A, Shichiri K, Tanaka A, Tanaka K, Inoue I. Genome-wide expression of azoospermia testes demonstrates a specific profile and implicates *ART3* in genetic susceptibility. *PLoS Genet* 4, e26, 2008.
2. Ikeda R, Yoshida K, Inoue I. Identification of FAZF as a novel BMP2-induced transcription factor during osteoblastic differentiation. **J Cell Biochem 101,147-154 2007.**
3. Systematic screening of lysyl oxidase-like (LOXL) family genes demonstrates that *LOXL2* is a susceptibility gene to intracranial aneurysms. Akagawa H, Narita A, Yamada H, Tajima A, Krischek B, Kasuya H, Hori T, Kubota M, Saeki N, Hata A, Mizutani M, Inoue I. **Hum Genet 121, 377-387, 2007.**
4. Increased expression of ephrin A1 in brain arteriovenous malformation.Sasahara A, Kasuya H, Akagawa H, Ujiie H, Kubo O, Sasaki T, Onda H, Sakamoto Y, Krischek B, Hori T, Inoue I. **Neurosurg Rev 30, 299-305, 2007.**
5. Pathopysiological role of endothelin in ectopic ossification of human spinal ligaments induced by mechanical stress.Iwasawa T, Iwasaki K, Sawada T, Okada A, Ueyama K, Motomura S, Harata S, Inoue I, Toh S, Furukawa K-I. **Calcif Tissue Int 79, 422-430, 2006.**
6. The alanine/threonine polymorphism of the alpha-1-antichymotrypsin (SERPINA3) gene and ruptured intracranial aneurysms in the Japanese population.Krischek B, Akagawa H, Tajima A, Narita A, Kasuya H, Hori T, Inoue I. **Cerebrovasc Dis 23, 46-49, 2006.**
7. The small heat shock protein alphaB-crystallin inhibits differentiation-induced caspase 3 activation and myogenic differentiation.Ikeda R, Yoshida K, Ushiyama M, Yamaguchi T, Iwashita K, Futagawa T, Shibayama Y, Oiso S, Takeda Y, Kariyazono H, Furukawa T, Nakamura K, Akiyama S, Inoue I, Yamada K.

Biol Pharm Bull 29, 815-819, 2006.

8. 井ノ上逸朗：多因子病の遺伝要因をどのように考えるべきか.「最新医学増刊 臨床遺伝子学'07−ゲノム科学の臨床へのインパクト−」(菅野純夫 編), 最新医学社, 62(9 増刊): 2124-2133; 2007.

9. 秋山康一、井ノ上逸朗：ゲノム全域 SNP タイピングでわかるゲノム構造変異,「骨・関節・靭帯」, アークメディア, 20(9): 845-854, 2007.

2) 学会発表

1. Yoshihara K, Sekine M, Tajima A, Inoue I, Tanaka K.: Two subclasses of advanced serous ovarian adenocarcinoma with different gene expression profilings by hierarchical clustering and non-negative matrix factorization. AACR Annual Meeting 2007, Los Angeles, USA. April 14-18, 2007.

H. 知的財産権の出願・登録状況

1.特許取得

なし。

2.実用新案登録

なし。

3.その他

なし

研究成果の刊行に関する一覧表

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
Okada H, Tajima A, Shichiri K, Tanaka A, Tanaka K, Inoue I	Genome-wide expression analyses of testes of non-obstructive azoospermia patients demonstrate a specific gene expression profile and implicate ART3 in genetic susceptibility.	PLoS Genet	4	1-14	2008
Takehiro Serikawa, Akira Kikuchi, Susumu Sugaya, Norio Suzuki, Hiroshi Kikuchi, and Kenichi Tanaka	In vitro and in vivo evaluation of novel cationic liposomes utilized for cancer gene therapy.	Journal of conto-rolled release	113	255-260	2006
Chiaki Banzai, Tetsro Yahata, Jun Sasahara, Katunori Kashima, Kazuyuki Fujita, Ken Nishikura, Yoichi Ajioka, Teiichi Motoyama, Kenichi tanaka	Advanced malignant rhabdoid tumor of the ovary effectively responding to chemotherapy : A case report and review of the literature.	Gynecologic Oncology	105	261-265	2007
Ikeda R, Yoshida K, Inoue I,	Identification of FAZF as a novel BMP2-induced transcription factor during osteoblastic diffe rentiation.	J Cell Biochem	101	147-154	2007
Akagawa H, Narita A, Yamada H, Tajima A, Krischek B, kasuya H. Hori T, Kubota, M, Saeki N, Hata A, Mizutani M, Inoue I	Systematic screening of lysyl oxidase-like(LOXL) family genes demonstrates that *LOXL2* is Suscepti-bility gene to intracranial aneurysms.	Hum Genet	121	377-387	2007
Sasahara A, kasuya H, Akagawa H, Ujiie H, Kubo O, Sasaki T. Onda H, Sakamoto Y, Krischek B, Hori T, Inoue I	Increased expression of ephrin Al in brain arteri-ovenous malformation.	Neurosurg Rev	30	299-305	2007

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
T Amikura, M Sekine, Y Hirai, S Fujimoto, I Kobayashi, T Fujii, I nagata, K Ushijima, K Obata, M Suzuki, M Yoshinaga, N Umesaki, S Satoh, T Enomoto, S Motoyama, K Nishino, K Haino, K tanaka	Mutational analysis of TP53 and p21 in familial and sporadic Ovarian cancer in Japan.	Gynecologic Oncology	100	365-371	2006
T Yokoo, K Takakuwa, I Ooki, A kikuchi, M Tamura, K Tanaka	Alteration of TH1 and TH2 cells by intracellular cytokine detection in Patients with unexplained recurrent abortion before and after immunotherapy with husbands mononuclear cells.	Fertility and Sterility	85	1452-1458	2006
S Kato, H Hanabusa, S Kaneko, K Takakuwa M Suzuki, N kuji, M Jinno, R Tanaka K Kojima, M Iwashita, Y Yosimura, K Tanaka	Complete removal of HIV-I RNA and proviral DNA from semen by the swim-up method:assisted reproduction technique Using spermatozoa free from HIV-1.	AIDS	20	967-973	2006
Serikawa T, Fugita K, Oite T, Tanaka K	Maintenance of in vitro granulose cell function by adenoviral mediated follicle stimulating hormone receptor gene transduction.	J Assist Reprod Genet	23	199-206	2006
Hiroaki Okada, Tetsuro Yahata, Jinhua Quan, Masayuki Sekine and Kenichi Tanaka	Association Between Single Nucleotide Polymorphisms Of Drug Resistance-associated Genes and Response To chemotherapy in Advanced Ovarian Cancer.	ANTICANCER RESERCH	26	2227-2232	2006

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
Takayuki Amikura, Yoichi Aoki, Chiaki Banzai, Tomokazu Yokoo, Nobumichi Nishikawa, Masayuki Sekine, Kenichi Tanaka	Metastatic choriocarcinoma successfully treated with paclitaxel and carboplatin after interstitial lung disease induced by EMA-CO	Gyinecologic Oncology	102	573-575	2006
Iwasawa T, Iwasaki K, Sawada T, Okada A, Ueyama K, Motomura S, Harata S, Inoue I, Toh S, Furukawa K-I	Pathopysiological role of endothelin in ectopic ossification of human spinal ligaments induced by mechanical stress.	Calcif Tissue Int	79	422-430	2006
Krischek B, Akagawa H, Tajima A, Narita A, Kasuya H, Hori T, Inoue I.	The alanine/threonine Polymorphism of the alpha-1-antichemotrypsin (SERPINA3) gene and ruptured intracranial aneurysms in the Japanese population .	Cerebrovasc Dis	23	46-49	2006
Ikeda R, Yoshida K, Ushiyama M, Yamaguchi T, Iwashita K, Futagawa T, Shibayama Y, Oiso S, Takeda Y, Kariyazono H, Furukawa T, Nakamura K. Akiyama S Inoue I , Yamada K	The small heat shock Protein alphaB-crystallin inhibits differentiation-induced caspase 3 activat-ion and myogenic differentiation.	Biol Pharm Bull	29	815-819	2006
Nakabayashi M, Ajioka Y, Lang I, Szentirmay Z, Kasler M, Nakadaira H, Yokoyama N, Watanabe G, Nishikura K, Wakai T, Shirai Y, Hatakeyama K, and Yamamoto M,	Genetic change of p53, K-ras, and microsatellite Instability in gallbladder Carciboma high-incidence areas of japan and Hungary	World J. Gastroenterol.	14	70-77	2008

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
Kajiura S, Yashiki T, Funaoka H, Ohkal Y, Nishikura K, Kanda T, Ajioka Y, Igarashi M, Hatakeyama K, Fujii H,	Establishment and characterization of monoclonal and polyclonal antibodies against human intestinal fatty asid-binding protein(I-FABP) using synthetic regional peptides and recombinanat I-FABP.	J. Imjunoassay Immunochem	29	19-41	2008
Kuwabara A, Okamoto H, Suda T, Ajioka Y, Hatakeyama K,	Clinicopathologic chara-cteristics of clinically relevant cytomegalovirus infection in inflammatory bowel disease.	J, Gasteroentrol	42	823-829	2007
Tozawa E, Akioka Y, Watanabe H, Nishkura K, Mukai G, Suda T, Kanoh T, Hatakeyama K.	Mucin expression, P53 over expression, and peritumoral lymphocytic infiltration of advanced colorectal carcinoma with mucus component: Is mucinous carcinoma a distinct histological entity?	Pathol Res Pract	203	567-574	2007
Tanabe K, Kand T, Kosugi S, ikeda Y, Makino S, Komukai S, Ohashi M, Suzuku T, Hatakeyama K.	Extranodal spreading of esophageal squamous cell arcinoma; clinicopathological Characteristics and prognostic impact.	World J. Surg	11	2192-8	2007
Korita PV, wakai T, Shirai Y, Sakata J, Tanizawa K, Cruz PV, Ajioka Y, Hatakeyama K	Intrahepatic lymphatic invasion independently predicts poor survival and recurrences after hepatectomy in patients with colorectal carcinoma liver metastasis.	Ann Surg Obcol	14	3472-80	2007

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
Yamamoto S, Sato Y, Oya H, Nakatsuka H, Kobayashi T, Hara Y, Watanabe T, Kurosaki I, Hatakeyama K.	Risk factors and prevention of biliary anastomotic complications in adult donor liver trans plaqntation.	World J. gastroenterol	13	4236-41	2007

書籍

著者氏名	論文タイトル名	書籍全体の編集者名	書籍名	出版社名	出版地	出版年	ページ
井ノ上逸朗	多因子病の遺伝要因をどのように考えるべきか	菅野純夫	最新医学増刊 臨床遺伝子学'07-ゲノム科学の臨床へのインパクト-	最新医学社	大阪	2007	2124-2133
秋山康一、井ノ上逸朗	ゲノム全域 SNP タイピングでわかるゲノム構造変異		骨・関節・靱帯	アークメディア	東京	2007	845-854

OPEN ACCESS Freely available online PLOS GENETICS

Genome-Wide Expression of Azoospermia Testes Demonstrates a Specific Profile and Implicates *ART3* in Genetic Susceptibility

Hiroyuki Okada[1☯], Atsushi Tajima[2☯], Kazuyoshi Shichiri[3], Atsushi Tanaka[4], Kenichi Tanaka[1], Ituro Inoue[2,5*]

1 Department of Obstetrics and Gynecology, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan, 2 Division of Molecular Life Science, School of Medicine, Tokai University, Isehara, Japan, 3 Department of Obstetrics and Gynecology, Tachikawa Hospital, Nagaoka, Japan, 4 St. Mother's Hospital, Kitakyushu, Japan, 5 Core Research for Evolutional Science and Technology, Japan Science and Technology Corporation, Kawaguchi, Japan

Infertility affects about one in six couples attempting pregnancy, with the man responsible in approximately half of the cases. Because the pathophysiology underlying azoospermia is not elucidated, most male infertility is diagnosed as idiopathic. Genome-wide gene expression analyses with microarray on testis specimens from 47 non-obstructive azoospermia (NOA) and 11 obstructive azoospermia (OA) patients were performed, and 2,611 transcripts that preferentially included genes relevant to gametogenesis and reproduction according to Gene Ontology classification were found to be differentially expressed. Using a set of 945 of the 2,611 transcripts without missing data, NOA was further categorized into three classes using the non-negative matrix factorization method. Two of the three subclasses were different from the OA group in Johnsen's score, FSH level, and/or LH level, while there were no significant differences between the other subclass and the OA group. In addition, the 52 genes showing high statistical difference between NOA subclasses ($p < 0.01$ with Tukey's *post hoc* test) were subjected to allelic association analyses to identify genetic susceptibilities. After two rounds of screening, SNPs of the ADP-ribosyltransferase 3 gene (*ART3*) were associated with NOA with highest significance with ART3-SNP25 (rs6836703; $p = 0.0025$) in 442 NOA patients and 475 fertile men. Haplotypes with five SNPs were constructed, and the most common haplotype was found to be under-represented in patients (NOA 26.6% versus control 35.3%, $p = 0.000073$). Individuals having the most common haplotype showed an elevated level of testosterone, suggesting a protective effect of the haplotype on spermatogenesis. Thus, genome-wide gene expression analyses were used to identify genes involved in the pathogenesis of NOA, and *ART3* was subsequently identified as a susceptibility gene for NOA. These findings clarify the molecular pathophysiology of NOA and suggest a novel therapeutic target in the treatment of NOA.

Citation: Okada H, Tajima A, Shichiri K, Tanaka A, Tanaka K, et al. (2008) Genome-wide expression of azoospermia testes demonstrates a specific profile and implicates ART3 in genetic susceptibility. PLoS Genet 4(2) e26. doi:10.1371/journal.pgen.0040026

Introduction

Spermatogenesis, a major function of mammalian testes, is complex and strictly regulated. While spermatogenesis is a maturation of germ cells, other cells including Sertoli, Leydig, and peritubular myoid cells also play important roles, and defects at any differentiation stage might result in infertility. Male infertility is estimated to affect about 5% of adult human males, but 75% of the cases are diagnosed as idiopathic because the molecular mechanisms underlying the defects have not been elucidated. In consequence, an estimated one in six couples experiences difficulty in conceiving a child despite advances in assisted reproductive technologies. Male-factor infertility constitutes about half of the cases, and a significant proportion of male infertility is accompanied by idiopathic azoospermia or severe oligozoospermia, which may well have potential genetic components. It is well-recognized that men with very low sperm counts (<1 million/ml), identified through an infertility clinic, have a higher incidence of Y-chromosome microdeletion (up to 17%) [1,2]. However, the genetic causalities of most cases of azoospermia are not known.

Global gene-expression profiling with microarray technologies has been applied with great promise to monitor biological phenomena and answer biological questions. Indeed, microarray technologies have been successfully used to identify biomarkers, disease subtypes, and mechanisms of toxicity. We applied microarray analysis to testis specimens from infertile individuals including patients with obstructive azoospermia (OA) and non-obstructive azoospermia (NOA [OMIM %606766]) to characterize NOA and to identify the specific pathophysiology and molecular pathways of the disease. In addition, we attempted to identify genetic susceptibility to NOA from genes differentially expressed in NOA testes.

Editor: Emmanouil T. Dermitzakis, The Wellcome Trust Sanger Institute, United Kingdom

Received July 31, 2007; **Accepted** December 13, 2007; **Published** February 8, 2008

Abbreviations: ART3, ADP-ribosyltransferase 3; AZF, azoospermia factor; coph, cophenetic correlation coefficient; Cy3, cyanine 3-CTP; Cy5, cyanine 5-CTP; EM, expectation–maximization; FDR, false discovery rate; FSH, follicle-stimulating hormone; GO, Gene Ontology; HC, hierarchical clustering; LD, linkage disequilibrium; LH, leutenizing hormone; MAF, minor allele frequency; MESA, microsurgical epididymal sperm aspiration; NMF, non-negative matrix factorization; NOA, non-obstructive azoospermia; OA, obstructive azoospermia; R^2, square of correlation coefficient; SNP, single nucleotide polymorphism; TESE, testicular sperm extraction

* To whom correspondence should be addressed. E-mail: ituro@is.icc.u-tokai.ac.jp

☯ These authors contributed equally to this work.

Author Summary

Worldwide, approximately 15% of couples attempting pregnancy meet with failure. Male factors are thought to be responsible in 20%–50% of all infertility cases. Azoospermia, the absence of sperm in the ejaculate due to defects in its production or delivery is common in male infertility. In this study, we focused on non-obstructive azoospermia (NOA) because the etiologies of obstructive azoospermia are well studied and distinct from those of NOA. Microdeletions of the Y chromosome are thus far the only genetic defects known to affect human spermatogenesis, but most cases of NOA are unsolved. Because NOA results from a variety of defects in the developmental stages of spermatogenesis, the stage-specific expressions of genes in the testes must be investigated. Thus, genome-wide gene expression analyses of testes of NOA can provide insight into the several etiologies and genetic susceptibilities of NOA. In the present study, we analyzed several differentially expressed genes in NOA subclasses and identified ART3 as a susceptibility gene for NOA.

Genes related to spermatogenesis and candidate genes for azoospermia have been surveyed in humans and mice, especially since gene targeting technology accelerated the identification of genes that play crucial roles in spermatogenesis [3]. Because spermatogenesis is a complex process including meiosis, a germ cell–specific event, gene expression profiles specific to the differentiation stage, clinically classified by the Johnsen's score, were examined to provide insight into the pathogenesis of azoospermia [4]. In the current study, we performed microarray analyses on biopsied testes obtained from 47 NOA patients at diverse clinical stages without prior selection and 11 OA patients. The 47 NOA samples showed a wide range of heterogeneity, including a series of impairments at the differentiation stage of spermatogenesis that so far have been evaluated mainly by pathological findings. Thus, classification of NOA at the transcriptome level is a necessary first step in elucidation of the molecular pathogenesis of NOA. To do this, we adopted the non-negative matrix factorization (NMF) method, an unsupervised classification algorithm developed for decomposing images that has been applied in various fields of science including bioinformatics because of its potential for providing insight into complex relationships in large data sets [5–7]. 47 of the NOA-samples were divided into three subclasses by the NMF method, and each class was associated with clinical features. 149 transcripts were identified as differentially expressed genes among the NOA subclasses according to a statistical criterion, and the features involved in spermatogenesis based on Gene Ontology classification were demonstrated.

The genetic causality of NOA most likely involves the expression level of a susceptibility gene, which might be detected by genome-wide gene expression analysis. While it is daunting to identify genetic susceptibility from 100-1000 differentially expressed genes, genetic susceptibility might more readily be identified from random genes differentially expressed with high significance rather than by investigating only genes in a specific biological pathway. Based on a well-defined statistical procedure, 52 candidate genes for NOA were catalogued by gene expression profile and screened for allelic association study in a total of 442 NOA patients and 475 fertile male controls. After gene-centric selection of SNPs, 191 SNPs of 42 candidate genes were initially evaluated for allelic association with NOA. After two rounds of screening, SNPs of the ADP-ribosyltransferase 3 (*ART3*) gene were found to be significantly associated with NOA, and five of these SNPs were selected for haplotype construction. The most common haplotype was significantly under-represented in the patients and may be protective. The functional impact of this haplotype was further investigated.

Results

Extraction of NOA-Related Gene Expression Profile

As shown in Figure 1A, the most notable difference in histological findings between NOA and OA testes was that the NOA patients exhibited, at varying degrees, incomplete sets of spermatogenic germ cells (spermatogonia, spermatocytes, spermatids, and spermatozoa) in the seminiferous tubules. In severe NOA patients, we could not even detect Sertoli cells, the major somatic cells of the seminiferous tubules, on histological examination (figure not shown), indicating clinical heterogeneity of NOA testes. In order to elucidate the molecular systems underlying NOA at the transcriptome level, it is important to extract genes reflecting the diversity of NOA phenotypes. For this purpose, we first compared global gene expression profiles in NOA testes to those of OA testes using the Agilent Human 1A(v2) Oligo Microarray system. We chose the 'standard reference design' in two-color microarray experiments as an experimental design for the expression analysis [8], where a single microarray was used to compare either NOA or OA to the testicular reference RNA as described in Materials and Methods (Figure 1B).

Of the 18,716 transcripts screened with the microarray, we obtained transcripts that showed a 2-fold mean expression difference between NOA and the reference, the NOA group; the OA group comprised transcripts showing less than 2-fold mean expression difference between OA and the reference (Figure 1B). Of the transcripts overlapping the two groups, 2,611 transcripts were found to be differentially expressed between NOA and OA testes after statistical filtering (based on lowess-normalized natural log[Cy5/Cy3], Bonferroni's corrected $p < 0.05$). This gene list, termed NOA-related target genes, comprised 902 elevated and 1,709 decreased transcripts in NOA testes. To characterize the gene list from the biological aspect, the 2,611 transcripts were subjected to functional clustering according to Gene Ontology (GO) classification for biological processes with GeneSpring software. We identified a total of 190 GO categories that were significantly ($p < 0.05$ without multiple testing correction) over-represented among the 2,611 transcripts. Table 1 shows the ten top-ranked GO categories in descending order of significance based on p-values with Fisher's exact test. It is notable that the GO categories involved in gametogenesis (GO:48232; 7283; 7276), reproduction (GO:19953; 3), and the cell cycle (GO:279; 51301; 7049; 7067) are significantly associated with the gene list. We further analyzed two separate subsets comprising 1,709 decreased (Figure 2, upper) and 902 elevated (Figure 2, lower) transcripts, based on their GO annotations. The top-ranked GO categories for NOA-related target genes are more similar to those of the 1,709 decreased transcripts than to those of the 902 elevated ones (Figure 2; Table 1). Thus, the predominant features may reflect spermatogenic defects common to NOA testes. On the

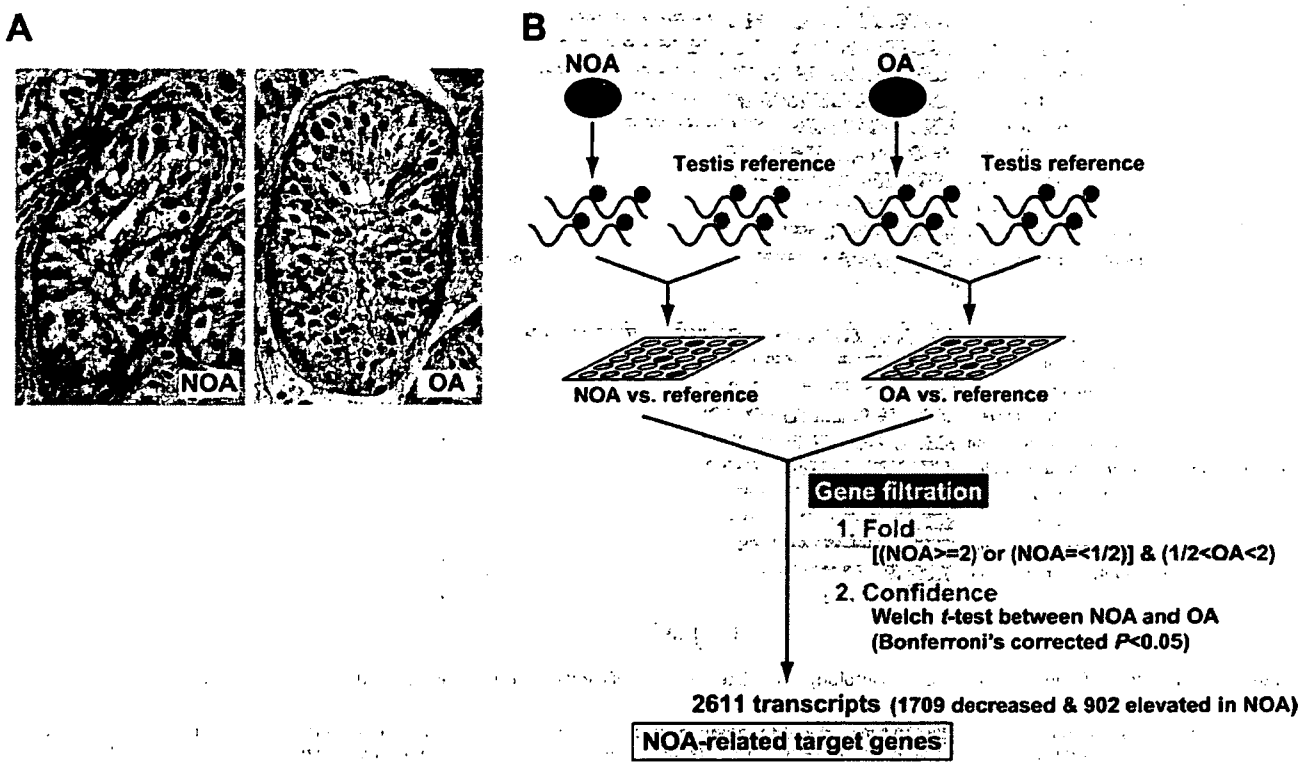


Figure 1. Experimental Design to Extract NOA-related Target Genes

(A) Representative testicular biopsies from NOA (left panel) and OA (right panel) patients. The seminiferous tubules in NOA testis show Sertoli cells only; testicular histology of the OA patient indicates the presence of germ cells at all stages of spermatogenesis.

(B) Strategy for discovering the NOA-specific expression profile, i.e., the NOA-related target genes. Compared with the expression level of reference RNA, the NOA group, with expression undergoing 2-fold or more mean change, was extracted; the OA group comprised transcripts with less than 2-fold mean expression change. Of the overlapping transcripts, only those with statistically significant difference between NOA and OA ($p < 0.05$ with Bonferroni's correction) were identified as differentially expressed.

doi:10.1371/journal.pgen.0040026.g001

Table 1. Top-Ranked Ten Categories of Gene Ontology Significantly Overrepresented among 2,611 Transcripts

GO Category[a]	Genes within GO Category		−logP[c]
	Number	Percent[b]	
Male gamete generation (GO:48232)	55	3.1	12.0
Spermatogenesis (GO:7283)	55	3.1	12.0
Sexual reproduction (GO:19953)	70	3.9	11.2
Gametogenesis (GO:7276)	59	3.3	10.5
Reproduction (GO:3)	75	4.2	8.3
M phase (GO:279)	57	3.2	6.7
Microtubule-based process (GO:7017)	42	2.4	6.5
Cell division (GO:51301)	48	2.7	6.5
Cell cycle (GO:7049)	144	8.1	6.4
Mitosis (GO:7067)	45	2.5	5.6

[a]All GO categories are from the subontology *biological process*.

[b]Percent denotes the percentage of coverage of NOA-related target genes. Of the 2,611 transcripts in the gene list, 1,784 are used for calculating the percentage of genes with a given GO annotation because the GO annotations regarding *biological process* for the others are not available.

[c]*p*-Value was determined by Fisher's exact test, comparing the observed percentage of NOA-related target genes with a given GO annotation to that of genes on the Agilent Human 1A(v2) microarray with the same GO annotation.

doi:10.1371/journal.pgen.0040026.t001

other hand, 902 transcripts elevated in NOA testes exhibited a distinct GO profile that included several GO categories involved in biosynthesis and metabolism in cytoplasm (Figure 2), implying an increase in cytoplasmic turnover rates such as steroid turnover in NOA testes.

Discovery of Three Molecular Subclasses of NOA Testes

To clarify heterogeneity of NOA testes at the transcriptome level, we further examined NOA-related target genes to identify NOA subclasses without prior classification with pathological features. We adopted the NMF algorithm coupled with a model selection method [6] to a complete dataset of 945 out of the 2,611 transcripts without missing values of signal intensities for a total of 47 NOA samples. Figure 3A shows reordered consensus matrices averaging 50 connective matrices generated for subclasses $K = 2$, 3, 4, and 5. Distinct patterns of block partitioning were observed at models with 2 and 3 subclasses ($K = 2$ and 3), whereas higher ranks ($K = 4$ and 5) made block partitioning indistinct. Thus, the NMF method predicts the existence of reproducible and robust subclasses of NOA samples for $K = 2$ and 3. This prediction was quantitatively supported by higher values of cophenetic correlation coefficients (coph) for NMF-clustered matrices. The NMF class assignment for $K = 3$ showed the highest coph value (coph = 0.993), indicating that three molecular subclasses, termed NOA1, NOA2, and NOA3, are

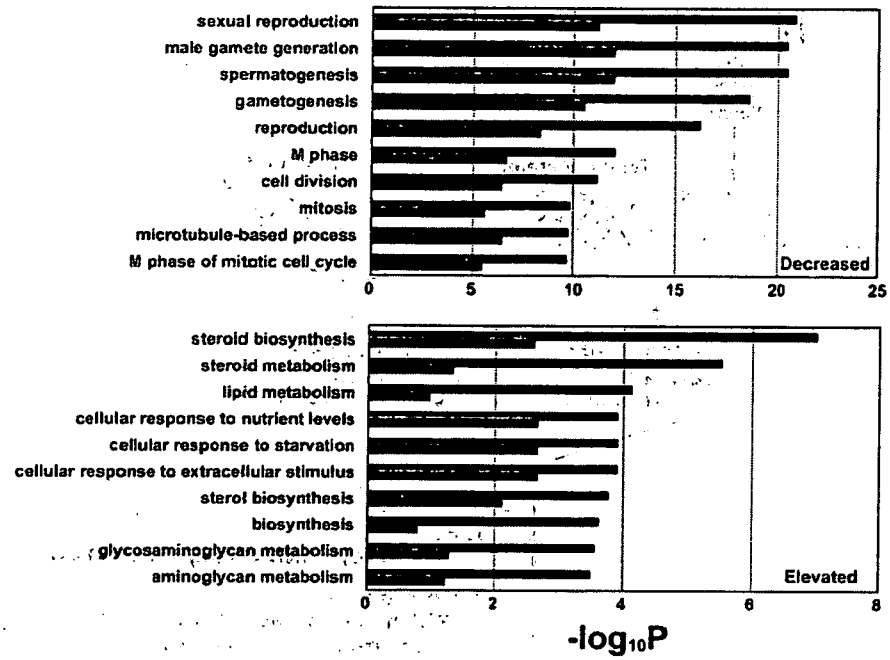


Figure 2. Comparison of Ten Top-ranked Gene Ontology (GO) Categories Significantly Over-Represented in the Two Subsets (1,709 Decreased, Upper Panel; 903 Elevated, Lower Panel) of NOA-related Target Genes

Blue and red bars represent *p*-values for the ten top-ranked GO categories over-represented in the decreased and elevated gene lists, respectively. Black bars show the corresponding *p*-value using 2,611 NOA-related target genes. *p*-Values were determined by Fisher's exact test using all of the GO-annotated genes on the Agilent Human 1A(v2) microarray as a background, as shown in Table 1. The *p*-values are expressed as the negative logarithm (base 10).

doi:10.1371/journal.pgen.0040026.g002

the most reasonable subclassification among 47 NOA samples. For comparative analysis of class discovery, a hierarchical clustering (HC) approach was applied to log-transformed normalized ratios for NOA-related target genes. As shown in Figure 3B, the HC dendrogram exhibited a clustering pattern similar to that of the NMF-based subclassification, as the three NMF-subclasses of NOA samples tended to form distinct clusters in the HC analysis. Thus, the HC clustering for NOA-related target genes appears to justify the three NMF-based subclasses of NOA samples.

To investigate the clinical features of the three NOA subclasses, we compared several clinical measures among the subclasses. The results obtained from statistical analyses in a total of four groups including the OA group are summarized in Table 2. We found significant differences in the three NOA-related clinical characteristics, testicular histological score (Johnsen's score, $p = 1.4 \times 10^{-6}$), serum FSH level ($p = 9.8 \times 10^{-4}$), and LH level ($p = 0.0051$) among the four groups using Kruskal-Wallis test, but there were no differences in age and serum testosterone level. *Post hoc* pairwise comparisons revealed that both the NOA1 and NOA2 groups exhibited low Johnsen's scores and high levels of serum FSH compared with the OA group (Table 2). In the NOA1 group, a high LH level ($p < 0.01$) also was found compared with the OA group. On the other hand, there were no significant differences in any of the parameters between the NOA3 and OA groups, as well as among three NOA subclasses in *post hoc* analysis. Elevations of serum FSH and LH concentrations often are observed in infertile patients with abnormal testicular histologies and are correlated, to some extent, with the severity of spermatogenic defects [9,10]. Testicular histologies of NOA and OA patients have been evaluated by the Johnsen's scores, ranging from 10 to 1 according to the presence or absence of spermatogenesis-related cell types (spermatozoa, spermatids, spermatocytes, spermatogonia, and Sertoli cells) in seminiferous tubules [11]. The NMF-based subclasses of testicular gene expression showed that the low score classes had heterogeneity (NOA1 and NOA2), presumably indicating the possibility of distinct spermatogenic defects at the molecular level that could not be detected by morphological examination.

Identification of Transcripts Differentially Expressed in the Three NOA Subclasses

Based on the three NOA subclasses, we conducted further statistical analyses to extract transcripts representing expression differences between NOA subclasses from the NOA-related target genes (Figure S1). 149 out of 2,611 transcripts showed significant differences ($p < 0.05$, Tukey's *post hoc* test) in testicular expression between the NOA subclasses, as summarized in Table S1. To characterize this gene list based on GO classification for biological processes, we examined which GO terms were highly associated with the 149 differentially expressed transcripts, relative to those for the NOA-related gene list (as shown in Table 1 and Figure 2). Figure 4 shows the 10 top-ranked GO categories for the 149 transcripts, using the 2,611 NOA-related target transcripts as a background set of genes for this GO analysis. Nine GO categories excluding gametogenesis appeared to be novel, indicating that highly significant enrichments of transcripts involved in DNA metabolism (GO:6259; 6325; 6323; 6281),

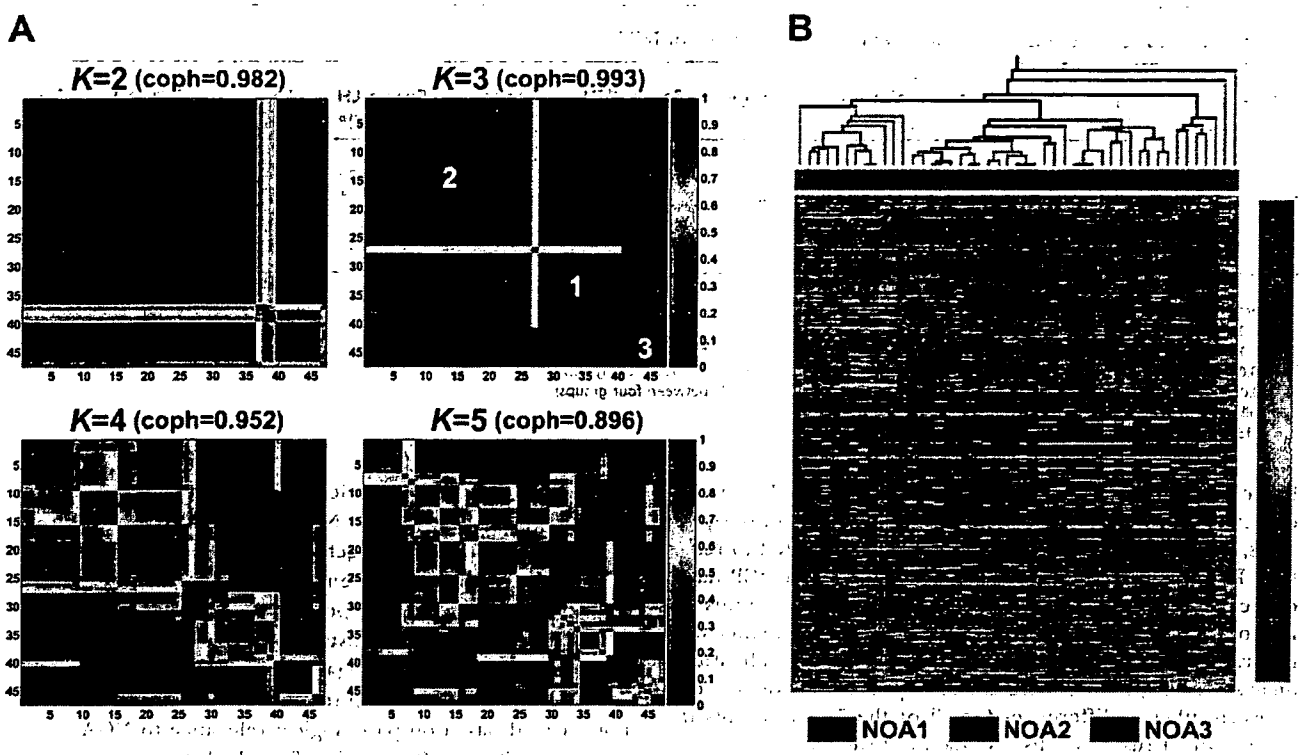


Figure 3. Non-Negative Matrix Factorization (NMF) and Hierarchical Clustering (HC)-based Subclassification of 47 NOA Samples

(A) Reordered consensus matrices averaging 50 connectivity matrices computed at $K = 2$–5 (as the number of subclasses modeled) for the NOA data set comprising of NOA-related target genes. NMF computation and model selection were performed according to Brunet et al. [6] as described in Materials and Methods. According to cophenetic correlation coefficients (coph) for NMF-clustered matrices, the NMF class assignment for $K = 3$ was the most robust.

(B) The HC method incorporated in GeneSpring also was used to classify NOA heterogeneity in testicular gene expression. To display correspondences of subclassification by the two methods, the NMF class assignments for $K = 3$ are shown color-coded; NOA1 (green), NOA2 (pink), and NOA3 (light purple).

doi:10.1371/journal.pgen.0040026.g003

chromosome organization and biogenesis (GO:51276; 7001), sex differentiation (GO: 7548), and response to endogenous stimulus (GO:9719; 6974) occurred after the extraction of 149 transcripts from the NOA-related target gene list (Figure 4). Other features of the 149 transcripts from the gene list (Table S1) were as follows: (1) a high frequency (24.2%) of sex chromosome-linked genes; (2) a high frequency (13.4%) of genes encoding cancer/testis antigens [12,13]; and (3) a moderate frequency (6.7%) of male infertility-related genes. Defect of these genes results in male infertility/subfertility in mice [3,14–16].

Twenty-five of the 149 transcripts showing differences in between-subclass expression displayed elevated expression in NOA, while the others (124 transcripts) had decreased expression (Table S1). The 25 NOA-elevated transcripts accounted only for differences in testicular expression between NOA1 and the other two subclasses, NOA2 and NOA3 (Figure S2; Table S1), suggesting testicular hyperactivity in NOA1 patients. For example, 3β-hydroxysteroid dehydrogenase, encoded by *HSD3B2* and *HSD3B1*, plays a crucial role in biosynthesis of testosterone in Leydig cells [17]. Expression levels of the two transcripts in the NOA1 subclass were higher than those in the NOA2 and NOA3 subclasses, and the expression difference between NOA1 and NOA3 was significant by Tukey's *post hoc* test (Figure S2; Table S1). As the NOA1 subclass showed significantly high LH and slightly low testosterone levels (Table 2), the elevated levels of the two transcripts may be explained by a compensation process for maintaining normal testosterone level. Thus, such enhanced steroidogenesis of the NOA1 subclass might favor, even if only slightly, testicular hyperactivity in NOA1 patients.

On the other hand, among the 124 NOA-decreased transcripts, most transcripts (118/124) showed expression differences between NOA3 and the other two subclasses (Figures S2–S4; Table S1). Expression levels of these transcripts in the NOA3 subclass were similar to those in testis reference RNA (Figures S2–S4), indicating that the NOA3 subclass has a mild defect in spermatogenesis. This notion is supported by the fact that the expression of *INHBB* encoding inhibin β subunit B in the NOA3 subclass is normal while NOA1 and NOA2 subclasses showed low levels, indicating that inhibin β may be a marker of testicular dysfunction, as previously reported [18].

Verification of Between-Subclass Differences in Testicular Expression by Quantitative Real-Time RT-PCR

To evaluate the appropriateness of microarray data on transcripts representing expression differences between NOA subclasses, we selected 53 with high significance ($p < 0.01$, Tukey's *post hoc* test, Figure S1 and Table S1) out of the 149 differentially expressed transcripts and subjected them to real-time RT-PCR analysis. Of the 53 transcripts, the highly

Table 2. Clinical Characteristics of Three Molecular Subclasses of NOA

Group	*n*	Age (Years)	Range	Johnsen's Score*	Range	Serum FSH (mIU/ml)*	Range	Serum LH (mIU/ml)*	Range	Serum T (ng/ml)	Range
OA	11	33.3 ± 8.5	25–57	7.9 ± 1.2	5.1–9	10.1 ± 9.3	3.6–31.4	4.5 ± 2.3	1.3–9.3	4.8 ± 1.7	3.4–7.0
NOA1	13	37.5 ± 6.0	27–52	2.0 ± 1.0[a]	1–4	34.5 ± 11.5[a]	19.0–53.3	13.6 ± 6.0[a]	5.2–20.8	2.8 ± 1.0	1.4–4.7
NOA2	27	34.4 ± 5.6	24–46	2.2 ± 1.1[b]	1–6	28.3 ± 6.4[c]	19.0–39.3	7.2 ± 2.6	2.4–13.0	3.7 ± 1.4	2.2–5.8
NOA3	7	33.0 ± 4.8	26–40	4.0 ± 1.6	2–6.5	22.7 ± 8.8	12.6–28.6[c]	5.9 ± 0.9	5.3–7.0	4.3 ± 2.7	2.0–7.3

The data are represented as mean ± standard deviation.
*$p < 0.01$ (Kruskal-Wallis test between four groups).
[a]$p < 0.01$, NOA1 versus OA (Scheffe's *posthoc* test on Johnsen's score, FSH and LH between four groups).
[b]$p < 0.01$, NOA2 versus OA (Scheffe's *posthoc* test on Johnsen's score, FSH and LH between four groups).
[c]$p < 0.05$, NOA2 versus OA (Scheffe's *posthoc* test on Johnsen's score, FSH and LH between four groups).
FSH, follicle-stimulating hormone; LH, luteinizing hormone; T, testosterone.
doi:10.1371/journal.pgen.0040026.t002

homologous VCX family genes, *VCX*, *VCX2*, and *VCX3A*, were detected with non-specific assay as a mixture of transcripts. Thus, 50 genes and one gene mixture were subjected to real-time RT-PCR. As shown in Figure S5, real-time RT-PCR data of the 51 transcripts were highly correlated with the results of microarray analysis, the squares of correlation coefficients (R^2) ranging from 0.40 (CT45–2) to 0.90 (GAJ). This validation analysis also provided statistically positive evidence on between-subclass differences for all of the 51 transcripts ($p < 0.05$ with Kruskal-Wallis test, data not shown).

Screening of Candidate Genes for Genetic Susceptibility for NOA

One approach to prioritizing candidate genes for genetic susceptibility underlying NOA is to adopt gene expression data from NOA tissues. Genes that show differences in expression level between NOA subclasses regardless of biological impact were selected based on the concept that polymorphic variation in gene expression among unrelated individuals is largely due to polymorphisms in DNA sequence [19,20]. 52 genes having statistical differences in expression ($p \leq 0.01$, Table S1) were regarded as candidates for allelic association with NOA. Despite the fact that these genes were not selected based on pathological relevance to NOA, genes such as *SYCP3*, *DAZL*, and *INHBB*, which were reported to function in spermatogenesis were included [21–23]. 191 single nucleotide polymorphisms (SNPs) of 42 genes were subjected to allelic association study with 190 NOA patients and 190 fertile men in the first round of screening. Ten genes

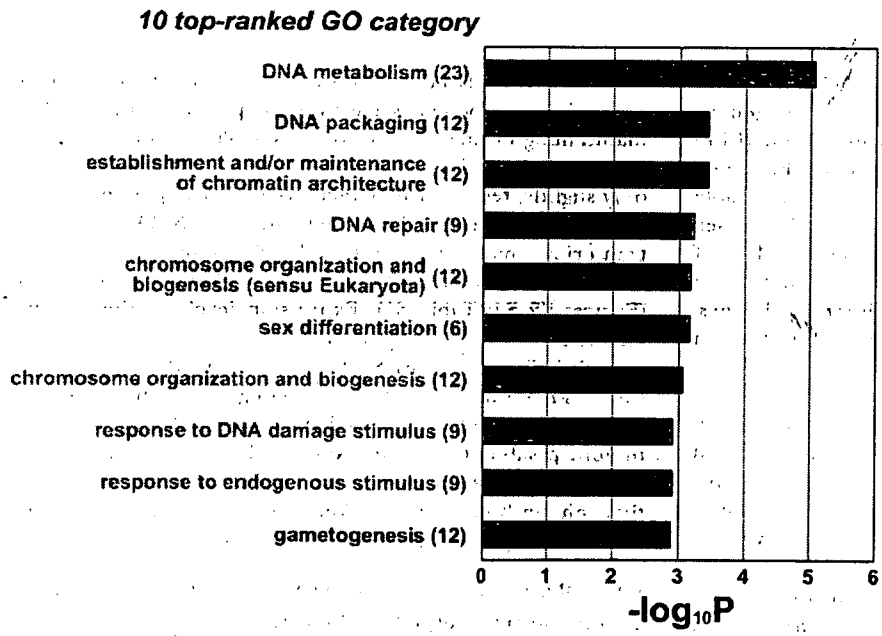


Figure 4. Significant Enrichments of Gene Ontology Categories in GO-based Profiling of 149 Differentially Expressed Transcripts

Grey bars represent *p*-values (expressed as the negative logarithm [base 10]) for the 10 top-ranked GO categories over-represented in the 149 differentially expressed transcripts, using the 2,611 NOA-related target transcripts as a background set of genes for the determination of *p*-values. The actual number of differentially expressed transcripts involved in each category is given in parentheses. In the process of extracting the 149 transcripts from the NOA-related target gene list, sets of genes involved in nine GO categories, marked in magenta, were more condensed because they were not found in the previous 10 top-ranked GO lists of NOA-related target genes (see Table 1 and Figure 2).
doi:10.1371/journal.pgen.0040026.g004

Table 3. Analyses of Allelic Association between 44 SNPs of Seven Candidate Genes and Japanese NOA in Second Round of Initial Screening (380 Cases versus 380 Controls)

Gene	Map	rs ID	Sequence Position[a]	Localization	Variation	MAF		χ^2	Nominal *p*
						Case	Control		
ART3	4q21.1	rs9995300	−2520	5′-upstream	A/G	40.7%	34.6%	5.60	0.018
		rs17001357	12048	Intron3	C/T	41.5%	46.4%	3.50	0.061
		rs11097230	19801	Intron3	A/G	32.9%	40.1%	7.79	0.0053
		rs6836703	34283	Intron11	G/A	41.4%	32.8%	11.7	0.00061*
		rs1128864	37726	Exon12-Non-synonymous	T/C	39.3%	33.2%	5.53	0.019
		rs6840007	43329	3′-downstream	A/T	34.5%	28.4%	6.16	0.013
LOC92196	2q24.1	rs4254463	−5901	5′-upstream	G/A	18.1%	16.7%	0.46	0.50
		rs908404	2538	Intron1	T/C	17.9%	16.1%	0.69	0.41
		rs9869	11757	Exon3-non-synonymous	C/T	19.7%	14.8%	5.85	0.016
		rs10016	20600	3′-UTR	G/A	9.8%	10.0%	0.00	0.99
NYD-SP20	17p13.3	rs3829957	−3785	5′-upstream	C/T	36.0%	36.2%	0.00	0.99
		rs2318035	12022	Intron5	A/G	36.9%	36.1%	0.09	0.77
		rs1488689	22797	Exon6-non-synonymous	A/G	34.8%	34.5%	0.01	0.94
		rs17822627	31572	Exon9-synonymous	T/C	35.6%	34.3%	0.23	0.63
		rs2318033	40646	3′-downstream	A/T	37.3%	36.4%	0.10	0.75
PAGE5	Xp11.21	rs2148982	2260	Intron3	G/A	30.9%	30.2%	0.01	0.92
		rs5913800	5332	3′-downstream	A/G	31.0%	29.8%	0.06	0.81
		rs5914276	8924	3′-downstream	C/G	31.4%	29.0%	0.40	0.53
		rs11091394	14424	3′-downstream	A/G	32.4%	28.2%	1.15	0.28
TEX14	17q22	rs686425	−7314	5′-upstream	G/A	47.1%	47.3%	0.00	0.99
		rs302874	1420	Intron1	C/T	47.3%	47.2%	0.00	0.99
		rs302865	12439	Intron1	C/T	46.7%	46.3%	0.01	0.93
		rs446613	19870	Intron1	A/G	47.6%	49.1%	0.27	0.60
		rs1631237	34721	Intron2	C/T	46.6%	47.3%	0.05	0.82
		rs302843	41430	Intron2	A/G	46.1%	49.0%	0.87	0.35
		rs2611782	51460	Intron2	C/T	47.5%	47.3%	0.00	0.99
		rs591200	63515	Intron2	C/T	42.8%	41.3%	0.25	0.62
		rs9898626	71197	Intron5	G/C	47.1%	46.4%	0.04	0.85
		rs302854	85804	Intron10	T/C	47.0%	48.0%	0.10	0.75
		rs8072873	100367	Intron15	G/C	24.6%	26.0%	0.30	0.58
		rs6503870	110398	Exon20-non-synonymous	T/C	46.8%	47.0%	0.00	0.95
		rs1267542	114726	Intron22	T/C	45.9%	49.1%	0.94	0.33
		rs3803751	119060	Intron24	T/C	23.5%	25.3%	0.61	0.43
		rs1267545	122507	Intron26	G/A	46.0%	48.1%	0.47	0.49
		rs1974586	128685	Intron29	C/T	23.5%	26.6%	1.67	0.20
		rs2333332	138362	3′-downstream	T/C	48.4%	49.0%	0.04	0.85
		rs714959	140815	3′-downstream	T/C	23.6%	22.6%	0.17	0.68
		rs12453459	145125	3′-downstream	C/T	49.0%	45.6%	0.90	0.34
TKTL1	Xq28	rs631	−8147	5′-upstream	G/A	23.4%	18.5%	2.37	0.12
		rs6655282	12986	Intron6	G/A	10.7%	6.8%	3.17	0.075
		rs766420	20834	Intron9	C/G	23.7%	22.7%	0.05	0.82
		rs2872817	24848	3′-UTR	A/G	28.5%	26.8%	0.18	0.67
		rs5945413	30181	3′-downstream	A/T	27.2%	27.0%	0.00	0.99
XAGE5	Xp11.22	rs4543711	5279	Intron4	A/G	4.9%	7.9%	2.30	0.13

[a]Nucleotide position from the first nucleotide of exon 1 of each gene.
*Statistically significant (corrected $p = 0.027$) based on Bonferroni-corrected *p*-value.
doi:10.1371/journal.pgen.0040026.t003

(*CTAG1B*, *LOC158812*, *LOC255313*, *MAGEA2*, *PEPP-2*, *TSPY1*, *TSPY2*, *VCX3A*, *VCY*, and *XAGE1*) were not analyzed because no gene-based SNPs with minor allele frequency (MAF) > 0.05 could be found. We identified seven genes (*ART3*, *LOC92196*, *NYD-SP20*, *PAGE5*, *TEX14*, *TKTL1*, and *XAGE5*) with at least one SNP showing a discrepancy in MAF of 5% or greater between cases and controls (Table S2). Forty-four SNPs in the seven genes were subjected to a second round of screening by increasing sample size (380 NOA patients and 380 fertile men). After the two rounds of screenings, only one SNP (rs6836703) of *ART3* (ADP-ribosyltransferase 3) was positively associated with NOA after Bonferroni's correction for multiple testing (Table 3; $\chi^2 = 11.7$, corrected $p = 0.027$).

Allelic Association Study with *ART3*

We focused on *ART3* based on the result of the two rounds of screenings, and identified 38 SNPs with MAF > 0.1 by database search or direct sequencing of the gene. 442 NOA patients (cases) and 475 fertile men (controls) were genotyped. Because we intended in this study to find a common genetic cause for NOA, patients with microdeletions of the Y chromosome at the azoospermia factor (AZF) locus, one of the major causes of NOA [1,2], were not excluded from the cases. However, to characterize the cases in regard to the AZF deletions, we examined the incidence of the deletions in a subset of the cases. Of the 442 NOA patients, 99 were examined by PCR-based screening. Fourteen (14.1 %) of the

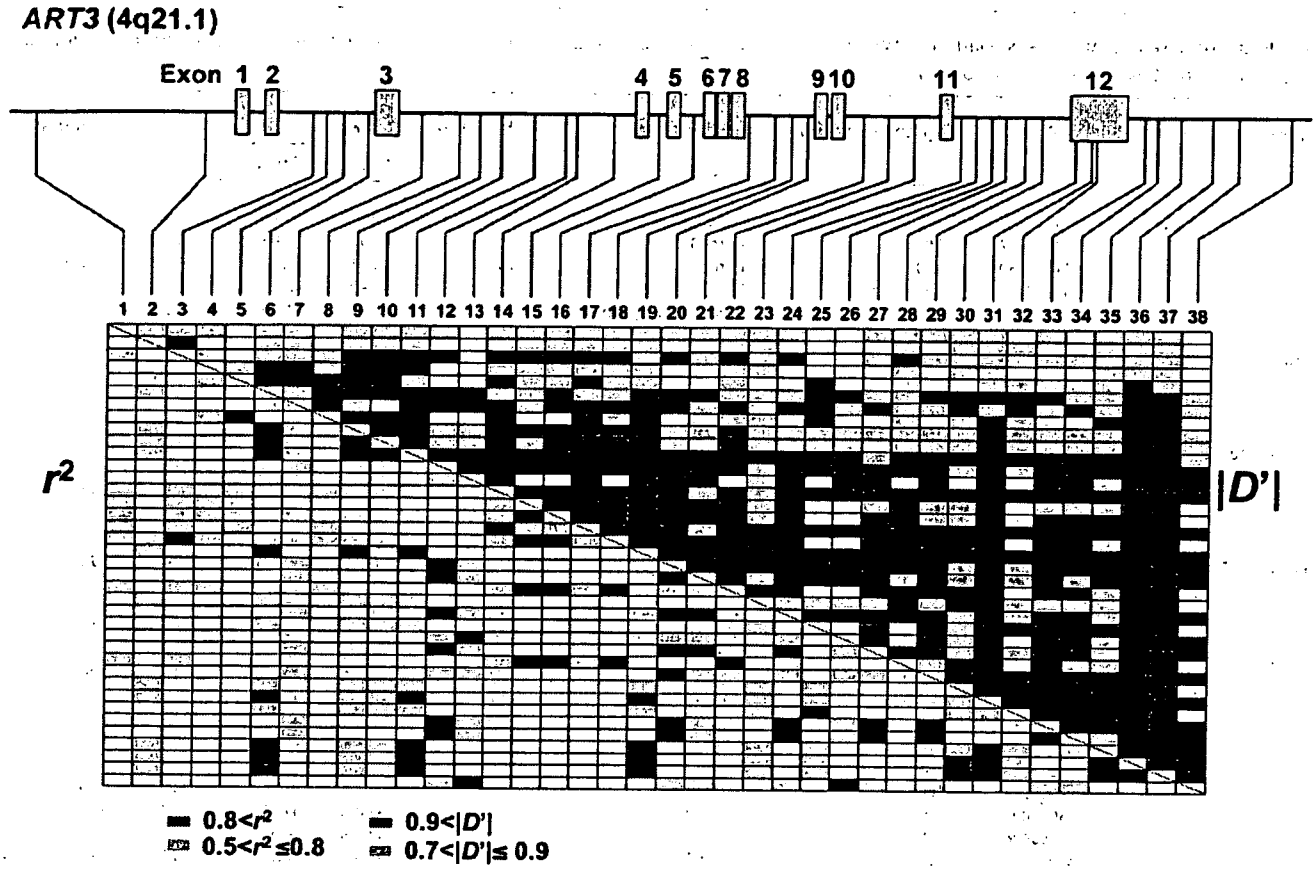


Figure 5. Linkage Disequilibrium Pattern of *ART3*

The gene structure together with the position of 38 SNPs is shown. Pairwise LD coefficients, D' and r^2, of controls were determined and expressed as a block structure. In the schematic block, red boxes indicate pairwise LD of $|D'| > 0.9$ or $r^2 > 0.8$ and pink boxes $0.9 \geq |D'| > 0.7$ or $0.8 \geq r^2 > 0.5$. Blank boxes represent $|D'| \leq 0.7$ or $r^2 \leq 0.5$.

doi:10.1371/journal.pgen.0040026.g005

99 cases examined showed the AZF deletions, and NOA patients with AZFc deletions were most frequent among the 14 cases (data not shown). The overall deletion frequency was comparable to those of other studies [1,2], in which the higher incidence of AZFc deletions also was observed. The clinical characteristics of patients with the AZF deletions did not differ from those of the other NOA patients (data not shown).

Linkage disequilibrium map showed that all of the SNPs of *ART3* were in near complete LD evaluated with D' statistic ($|D'| > 0.7$) in both cases and controls (only controls are displayed in Figure 5). None of the SNPs in the controls showed deviation from Hardy-Weinberg's equilibrium at a threshold of $p < 0.01$ (data not shown). As shown in Table 4, SNPs showing positive associations based on nominal p-values were widely distributed throughout *ART3*. The most significant association was observed with ART3-SNP25 (rs6836703) located in intron 11 of *ART3* ($\chi^2 = 9.16$, nominal $p = 0.0025$, odds ratio [95% CI] = 1.34 [1.11–1.63]). We applied the permutation method for adjustment of multiple testing to avoid a false positive result [24]. A total of four SNPs including ART3-SNP25 met the empirical significance level of $p < 0.05$ (Table 4).

For the haplotype-based association study, we first selected five SNPs (ART3-SNP1, 5, 23, 25, and 28) as tag SNPs captured through LD in *ART3* from 15 SNPs with nominal $p < 0.05$ at a threshold of $r^2 \geq 0.8$ with Tagger software [25]. Haplotype frequencies were inferred using an expectation-maximization (EM) algorithm. After excluding rare haplotypes (frequency < 0.01), association of ART3 haplotypes with NOA was examined in 442 cases and 475 controls. Haplotype H1, the most common haplotype in controls, was under-represented in cases with significance (Figure 6; 26.6% in cases and 35.3 % in controls; $\chi^2 = 15.7$, df = 1, nominal $p = 0.000073$), indicating a protective impact of haplotype H1. After Bonferroni's correction for multiple testing, a protective effect of haplotype H1 was still significant (corrected $p = 0.00080$). Other haplotypes showed no significant difference in frequencies between cases and controls (Figure 6). We also applied a Bayesian algorithm for phasing haplotypes with PHASE version 2.1.1 [26,27]. Regardless of haplotype-phasing methods, haplotype H1 was the most frequent in controls (26.4% in cases and 35.0% in controls), and a significant difference in haplotype H1 frequency between cases and controls was observed (permutation $p < 0.0001$ in global comparison, generated after 10,000 iterations).

Clinical Relevance of the Haplotype Associations

The functional relevance of haplotype H1 in comparison with the clinical data was then explored. Diplotype was

Table 4. Allelic Association of 38 SNPs in *ART3* with NOA in Japanese Population (442 NOA Patients versus 475 Controls)

Number	rs ID	Sequence Position[a]	Localization	Variation	MAF		χ^2	Nominal *p*	Permutation *p*[b]
					Case	Control			
ART3–1	rs13111494	−10376	5′-upstream	T/C	37.7%	43.4%	3.96	0.047	0.36
ART3–2	rs9995300	−2520	5′-upstream	A/G	38.7%	35.8%	1.66	0.20	0.81
ART3–3	rs4859609	3333	intron2	A/G	39.3%	43.2%	2.84	0.092	0.55
ART3–4	rs7666159	3941	intron2	T/C	51.5%	48.7%	1.40	0.24	0.87
ART3–5	rs4859611	5562	intron2	T/C	43.9%	48.8%	4.42	0.035	0.29
ART3–6	rs10007524	7070	intron2	G/A	34.6%	30.7%	3.14	0.077	0.49
ART3–7	rs4859612	9950	intron3	T/G	16.8%	17.8%	0.29	0.59	1
ART3–8	rs17001357	12048	intron3	C/T	40.6%	45.9%	5.13	0.024	0.22
ART3–9	rs4308383	12986	intron3	C/T	35.8%	31.2%	4.24	0.040	0.32
ART3–10	rs17001364	13478	intron3	T/C	36.4%	31.2%	5.47	0.019	0.19
ART3–11	rs7675618	15549	intron3	G/A	35.0%	30.3%	4.33	0.038	0.31
ART3–12	rs4859422	16736	intron3	G/A	23.2%	25.1%	0.88	0.35	0.96
ART3–13	rs4859614	17327	intron3	G/A	32.8%	31.2%	0.51	0.48	1
ART3–14	**rs11097230**	**19801**	**intron3**	**A/G**	**32.5%**	**39.3%**	**8.73**	**0.0031**	**0.040**
ART3–15	rs6829592	23717	intron4	G/A	43.3%	46.0%	1.32	0.25	0.89
ART3–16	rs13131187	25213	intron5	A/G	42.3%	45.1%	1.46	0.23	0.86
ART3–17	**rs17001385**	**26422**	**intron8**	**C/G**	**32.5%**	**39.2%**	**8.84**	**0.0030**	**0.038**
ART3–18	rs12331871	27635	intron8	T/G	39.7%	43.5%	2.57	0.11	0.61
ART3–19	rs17001390	28316	intron8	C/T	34.3%	30.2%	3.49	0.062	0.43
ART3–20	rs12510869	28670	intron8	A/G	23.2%	24.4%	0.36	0.55	1
ART3–21	rs13130116	29976	intron10	C/T	22.6%	24.9%	1.28	0.26	0.90
ART3–22	rs9307076	31256	intron10	G/A	40.3%	43.5%	1.92	0.17	0.75
ART3–23	rs4599438	32278	intron10	A/G	37.5%	31.7%	6.71	0.010	0.11
ART3–24	rs17001409	33549	intron11	T/C	23.8%	25.1%	0.39	0.53	1
ART3–25	**rs6836703**	**34283**	**intron11**	**G/A**	**41.2%**	**34.3%**	**9.16**	**0.0025**	**0.034**
ART3–26	rs4241584	35127	intron11	C/T	30.3%	29.5%	0.13	0.72	1
ART3–27	rs4859423	35158	intron11	T/C	23.7%	25.3%	0.67	0.41	0.99
ART3–28	rs4241586	35506	intron11	T/C	39.7%	44.5%	4.14	0.042	0.33
ART3–29	rs17001416	35935	intron11	G/T	21.6%	22.5%	0.20	0.66	1
ART3–30	rs1128864	37726	exon12-non-synonymous	T/C	38.2%	33.8%	3.72	0.054	0.39
ART3–31	New	37857	exon12–3′-UTR	G/A	34.1%	29.4%	4.58	0.032	0.28
ART3–32	**rs14773**	**37861**	**exon12–3′-UTR**	**A/C**	**45.0%**	**38.1%**	**8.96**	**0.0028**	**0.036**
ART3–33	rs7689378	38491	3′-downstream	A/G	21.7%	23.4%	0.67	0.41	0.99
ART3–34	rs13141802	38730	3′-downstream	G/C	22.3%	24.2%	0.86	0.35	0.97
ART3–35	rs10654	40031	3′-downstream	T/A	34.8%	30.5%	3.64	0.056	0.41
ART3–36	rs7675107	41918	3′-downstream	A/G	33.6%	29.1%	4.19	0.041	0.32
ART3–37	rs6840007	43329	3′-downstream	A/T	34.3%	29.7%	4.42	0.036	0.30
ART3–38	rs4538520	49092	3′-downstream	C/T	33.1%	31.7%	0.39	0.53	1

SNPs in bold show statistical significance and are subjected to haplotype analysis as shown in Figure 6.
[a]Nucleotide position from the first nucleotide of exon 1.
[b]Permutation *p*-values generated by 10,000 iterations. SNPs in bold show statistical significance based on the permutation *p*-values.
doi:10.1371/journal.pgen.0040026.t004

inferred with EM algorithm, and three categories (code 0, 1, and 2) were defined by the number of haplotype H1 carried without counting the other haplotypes, and nonparametric analysis of variance test with clinical data was performed. Serum levels of hormones (LH, FSH, and testosterone), other biochemical and pathophysiological markers, and Johnsen's score were analyzed by Kruskal-Wallis test with a Bonferroni/Dunn *post hoc* test between the three diplogroups. Serum testosterone levels were significantly different among the three groups (Figure 7; df = 2, $p = 0.0093$), but there were no significant differences in other clinical markers. *Post hoc* pairwise comparisons revealed that serum testosterone levels were significantly higher in the subgroup having two copies of haplotype H1 than in a subgroup with one or no haplotype H1 ($p = 0.0064$ or $p = 0.0004$, respectively, Figure 7). PHASE-inferred individual diplotypes also revealed a similar correlation between diplogroups of haplotype H1 and serum testosterone levels (data not shown).

ART3 Protein Localization in Azoospermic Testis

ART3 protein expression in azoospermic testes was examined by immunohistochemical analysis. As shown in Figure 8, specific staining of ART3 protein was predominantly observed in spermatocytes in OA testes (Figure 8C–8E) as well as in normal testes from individuals of accidental sudden-death (Figure 8A and 8B). Staining was not observed in other stages of undifferentiated germ cells or Sertoli cells in the seminiferous tubules, or the interstitial tissues such as Leydig cells. On the other hand, we did not detect any ART3 protein in NOA testes with Johnsen's scores ranging from 2 to 3, which showed no spermatocytes, spermatids, or spermatozoa in the seminiferous tubules ($n = 12$ samples; Figure 8F–8H). There was no marked difference in testicular ART3