A Common Mutation in *DEFB126* Causes Impaired Sperm Function and Subfertility

Summary: A frameshift mutation in *DEFB126* has a high allele frequency in multiple human populations, causes significant alteration in the sperm surface composition, significantly impairs the sperms ability to penetrate mucus-like gels, and in a prospective cohort study, leads to subfertility.

Theodore L. Tollner[1*], Scott A. Venners[2*], Edward J. Hollox[3], Ashley I. Yudin[4], Xue Liu[5], Genfu Tang[5], Houxun Xing[5], Robert J. Kays[6], Tsang Lau[6], James W. Overstreet[4], Xiping Xu[7], Charles L. Bevins[6#] and Gary N. Cherr[8#]

1 Department of Obstetrics and Gynecology, School of Medicine, University of California, Davis CA, USA
2 Faculty of Health Sciences, Simon Fraser University, Burnaby, Canada
3 Department of Genetics, University of Leicester, UK
4 Center for Health and the Environment, University of California, Davis CA, USA
5 Institute for Biomedicine, Anhui Medical University, Hefei, China
6 Department of Microbiology & Immunology, School of Medicine, University of California, Davis CA, USA
7 University of Illinois at Chicago, School of Public Health, Chicago, IL, USA
8 Departments of Environmental Toxicology and Nutrition, Bodega Marine Laboratory, University of California, Davis CA, USA

*, # These authors contributed equally to this work.

Address Correspondence To:

Gary N Cherr
Bodega Marine Laboratory
University of California Davis
PO Box 247
Bodega Bay, CA 94923
gncherr@ucdavis.edu

Abstract

A glycosylated polypeptide, β-defensin 126 (DEFB126), derived from the epididymis and adsorbed onto the sperm surface, has been implicated in immunoprotection and efficient movement of sperm in mucosal fluids of the female reproductive tract. Here, we report a sequence variant in *DEFB126* that has a 2-nucleotide deletion in the open reading frame, which generates a non-stop mRNA. The allele frequency of this variant sequence is high in both a European (0.47) and a Chinese (0.45) population cohort. Binding of the *Agaricus bisporus* lectin to the sperm surface glycocalyx was significantly lower in men with the homozygous variant (del/del) genotype than in those with either a del/wt or wt/wt genotype, suggesting an altered sperm glycocalyx with fewer O-linked oligosaccharides in del/del men. Moreover, sperm from the del/del donors exhibited an 84% reduction in the rate of penetration of a hyaluronic acid (HA) gel, a surrogate for cervical mucus, compared to the other genotypes. This reduction in sperm performance in HA gels was not a result of decreased progressive motility (average curvilinear velocity) or morphological deficits. However, *DEFB126* genotype and lectin binding were highly correlated with performance in the penetration assays. In a prospective cohort study of newly married couples who were trying to conceive by natural means, couples were less likely to become pregnant and took longer to achieve a live birth if the male partner was homozygous for the variant sequence. This common sequence variation in *DEFB126*, and its apparent cause of impaired reproductive function, provides an opportunity to better understand, clinically evaluate, and possibly treat human infertility.

Introduction

According to the World Health Organization (WHO), human infertility is defined as the inability for a couple to conceive after 1 year of unprotected sexual intercourse (1, 2). By this definition, the prevalence of infertility in many countries of the world is approximately 13 – 14% (3). In approximately half of infertile couples, the cause of infertility lies with the male partner. Infertility in males is usually evaluated by analysis of semen quality, as assessed by sperm count in the ejaculate, the percentage of motile sperm and the percentage of sperm with normal morphology. However, other than very low numbers of sperm, none of these measures alone is strongly diagnostic of infertility (4). Although improved estimates of male fecundity can be achieved by evaluating combinations of these and other semen factors including sperm velocity and tests of hyperosmotic swelling (5, 6), infertility is unexplained in approximately 17% of infertile couples (7). In these cases, no reproductive function abnormalities can be established on the basis of currently available assessments.

The elaborate glycocalyx of sperm is a conserved feature of epididymal maturation in mammals, yet how the sperm glycocalyx contributes to male fertility is not well understood (8). The dense carbohydrate coat provides protection for sperm during transit in the epididymis and female reproductive tract (8), and also assists with other key functions, including attachment of sperm to oviductal epithelium, regulation of capacitation, and sperm-egg interaction (9). Consistent with a proposed functional role of this sperm coat, differences in lectin labeling of the sperm glycocalyx between fertile and subfertile males have been detected in diverse species including fowl (10), livestock (11), and humans (12, 13) but the biochemical underpinnings of these observations remain elusive.

Recent studies with the sperm surface protein β-defensin 126 (DEFB126) provide some insight into the importance of glycocalyx structure to the fertilizing potential of sperm. In the innate immune system, defensins are expressed in phagocytic leukocytes and at surface epithelia, where these peptides serve as key antimicrobial effector molecules (14). The epididymis is another site of defensin expression, but in this case their function appears linked with reproductive physiology. In macaques, epididymal DEFB126 is a highly sialylated glycopeptide adsorbed to the surface of sperm during transit through the epididymis (15) and becomes a major constituent of the sperm glycocalyx (16). The deduced amino acid sequence of human DEFB126 predicts 20 potential sites for O-linked glycosylation in the 52-amino acid tail extending from the carboxyl terminus of the defensin peptide core. This β-defensin is retained on macaque sperm as they advance into the upper female reproductive tract (17), where it imparts an immuno-protective coat (18), facilitates sperm penetration of cervical mucus (19), and mediates sperm binding to oviductal epithelium (17). Similar structure and function is apparent in the mouse ortholog (βdef22), but this molecule has been less studied (20). From an even broader perspective, cysteine-rich defensin-like peptides appear to be integral to reproductive success in invertebrates, as well as in the plant kingdom (21, 22, 23).

Here, we present evidence that a high percentage of men carry a functional polymorphism in the gene encoding DEFB126 and evaluate its role in conception.

Results

Common polymorphism in human *DEFB126* predicts a non-stop mRNA.

The *DEFB126* gene is at the subtelomeric end of 20p13 in humans and does not show copy number variation (24). The encoded protein, DEFB126, has the canonical six-cysteine

array characteristic of defensins in the β- defensin subgroup (16), but is different from most other defensins because it has an extended C-terminal tail of 52 amino acids (25). This domain is the site of O-linked glycosylation. Orthologs of this defensin with highly similar sequences exist in non-human primates (Figure 1) and in mice (where it is named βdef22) (20). During the course of cloning human *DEFB126* cDNA, we identified a sequence variation (Figure 2A and 2B), and the variant was confirmed in the NCBI database (accession AK22598) and dbSNP (rs11468374). This *DEFB126* sequence polymorphism is common in several population cohorts (allele frequency 0.44-0.61) (Table 1) and the distribution of allele frequencies approximates Hardy-Weinberg equilibrium in these populations. The polymorphism is a two-nucleotide omission that results in a reading frame shift and generates a non-stop mRNA. Published analyses of other genes with mutations that result in mRNAs lacking in-frame stop codons found that the aberrant mRNAs were less abundant than the corresponding wild-type mRNAs (26, 27), and the reduced levels were attributed to a so-called nonstop mRNA decay surveillance mechanism (28, 29, 30). In addition, the translation of mRNA lacking in-frame stop codons is impaired (31). Consistent with these findings, *DEFB126* mRNA expression was lower in epididymal tissue with a del/del than a wt/wt genotype (Figure 2C).

Sperm from del/del donors have reduced surface glycosylation associated with O-linkages.

Twenty-one semen donors recruited for sperm function studies were genotyped for the *DEFB126* sequence variant. The frequency of the variant allele was 0.54 in this pool of donors. Nineteen of the donors received semen evaluations. There was no association of *DEFB126* genotype with any of the measured parameters of the semen analysis (semen volume, sperm density, percent sperm motility, and total motile count; Table 2). The sperm from these donors

were labeled with the *Agaricus bisporus* (ABA) lectin, which selectively binds O-linked galactose-GalNAC glycans. Sperm from donors with a *DEFB126* variant (del/del) genotype consistently showed lower ABA-associated fluorescence than the other two genotypes (Figure 3). Quantification of sperm fluorescence indicated that the differences in ABA labeling observed with genotype were significant ($p = 0.0006$; Figure 4D). The marked reduction in binding sites for ABA suggests that the glycocalyx of sperm from men with the del/del genotype lacks most of its O-linked oligosaccharides.

Sperm from del/del donors exhibit reduced HA gel penetration ability

We evaluated the ability of human sperm to penetrate a viscous hyaluronic acid (HA) gel, a surrogate for cervical mucus (Figure 4A). Because human cervical mucus is of limited availability and high variability, HA gels are used to simulate cervical mucus for *in vitro* analysis of sperm function (32). Although cervical mucus has complex biophysical properties that are derived from at least 5 distinct mucin molecules produced at the cervix (33, 34), gels prepared from HA share some of the properties of mucus, especially with respect to viscosity and charge (35). In addition, HA gels resemble cervical mucus in their penetrability by human sperm (36, 37, 38) and, in the macaque, subtle manifestations of cryodamage to frozen-thawed sperm were reflected equally in penetration tests using either HA or cervical mucus (39). Sperm from donors that were homozygous for the *DEFB126* polymorphism (del/del; n=6) exhibited reduced ability to penetrate HA, compared to sperm from men with either of the other two genotypes (wt/wt; n=6 or wt/del; n=4; $p = 0.008$). Both *DEFB126* genotype (Figure 4A) and lectin labeling intensity (Figure 4D) correlated with the results of the HA penetration assay (Figure 5). In contrast, there is no association of DEFB126 genotype with either VCL or morphology (Figure

4B and 4C), suggesting that the changes in HA penetration observed with genotype are not associated with changes in these classical assessments of sperm. There were no associations of *DEFB126* genotype with any category of abnormal sperm forms or with any of the sperm motion parameters measured by computer-assisted sperm analysis (Table 2), in spite of the fact that sperm morphology and progressive motility are regarded as the best available quantitative parameters for sperm function in HA gel and cervical mucus penetration assays (40, 41, 42, 43). In contrast, lectin-labeling intensity of sperm correlated well with the efficiency of sperm penetration of HA gels ($r = 0.800$; $p = 0.003$; Figure 5).

DEFB126 gene polymorphism in men is associated with reduced fertility

We examined the prevalence of the sequence variant in men and its association with fertility in a population-based, prospective cohort study of newly married couples in Anhui Province, China. Couples had no history of infertility and began attempting to conceive by natural means soon after enrollment (between July, 2003 and February, 2005). The median time from enrollment to follow-up was 22 months (minimum 21 months, maximum 40 months). Pregnancy and birth outcome were defined by self-report at the time of the follow-up interview, and the date of birth was verified with birth certificates.

Out of 812 men from couples at baseline who provided a blood sample, we randomly selected 664 for *DEFB126* genotyping. We successfully genotyped 638 men for whom we obtained DNA of sufficient quality and quantity. The *DEFB126* genotype frequencies for the 638 men were 187 (29%) homozygous wt, 328 (51%) heterozygous wt/del, and 123 (19%) homozygous del (Table 1), which approximated Hardy-Weinberg equilibrium (using Pearson goodness-of-fit test) (44).

The data for couples with male *DEFB126* genotype information were analyzed by logistic regression for the relative odds of self-reported pregnancy at follow-up according to the male's *DEFB126* genotype. Couples were excluded from this analysis for history of female-related factors of infertility including cervical polyps, uterine myoma, ovarian tumor, pelvic inflammation and smoking (n=38). Forty-eight couples were excluded who had used oral contraceptives or an IUD within the year prior to enrollment. Data were missing for another 43 couples who were lost to follow-up, leaving 509 couples for analysis of the odds of pregnancy. The mean (standard deviation) ages in this group were 25.8 (2.6) for men and 23.4 (2.2) years for women. The means (standard deviations) for months of follow-up during which a pregnancy could have occurred were 26 (6), 25 (5) and 25 (5) for the wt/wt, wt/del and del/del groups respectively. Our analysis showed that the odds of pregnancy (ratio of those who became pregnant to those who did not) among couples in which males had the del/del genotype were 60% of those observed for couples in which males had either wt/wt or del/wt genotypes (OR=0.6, p=0.029, Table 3). These data reveal a statistically significant decrease in fertility for males with *DEFB126* del/del genotype, when considering odds of pregnancy.

We also analyzed the time to live birth using Cox proportional hazards models. For this analysis, we excluded an additional 29 couples who achieved pregnancy, but who had spontaneous or induced abortion, leaving 480 couples for analysis of time to live birth. The mean (standard deviation) ages in this group were 25.8 (2.6) for men and 23.3 (2.3) years for women. Among couples in which males had the del/del genotype, the average (standard deviation) time from enrollment to the live birth of a child (or to the end of follow-up if there was no birth) was 17.4 (7.4) compared to 15.7 (7.3) months for couples in which males had either wt/wt or del/wt genotypes. Using proportional-hazards regression modeling, we

calculated for couples in which males had the del/del genotype a probability of live birth per month that was 70% of that determined for couples in which males had either wt/wt or wt/del genotypes (OR 0.7, p= 0.026, Table 4). Those who at follow-up had not yet become pregnant (n=109), or who were pregnant but had not yet given birth (n=40), contributed right-censored data for this analysis, meaning the follow-up times without events of live birth were included in the calculations of the probability per month of live birth. These data reveal a statistically significant decrease in fertility for males with *DEFB126* del/del genotype, when considering time to live birth.

Discussion

Studies in non-human primates have shown that DEFB126 is a major component of the sperm surface glycocalyx and is important for normal sperm function, including efficient sperm trafficking in the female reproductive tract (15, 16, 17, 18, 19, 45, 46). We now report that a mutation in human *DEFB126* is common in each of several disparate population cohorts interrogated (allele frequency 0.44-0.61). The genetic variant is a frameshift 2-nucleotide deletion, creating a non-stop mutation in the mRNA. Men who are homozygous for the deletion mutation produce sperm that have a deficit in surface O-linked oligosaccharides and exhibit difficulty penetrating HA gels *in vitro*. Yet, with respect to common measures of semen quality (sperm density, percentage of motile sperm, sperm progressive motility, and sperm morphology), these men appear normal according to WHO criteria and resemble men who posses the wt *DEFB126* allele. Findings from our analysis of a population-based prospective cohort show that men with the del/del genotype are significantly less fertile than men who carry the wt allele. Therefore, the *DEFB126* genotype could be a useful parameter in evaluation of male infertility.

DEFB126 is expressed by cells of the epididymis and deposited on the sperm surface as they mature and transit through this tissue. In men who are homozygous for the mutant allele, this maturation event results in an unusual (and perhaps unprecedented) situation in which the genetic variant that affects the expression of a protein of one tissue (the epididymis) alters the surface properties and function of a cell (sperm) with a completely different tissue origin (the testis). Prior studies demonstrated that removal of DEFB126 from surface of macaque sperm reduces sperm penetration of cervical mucus; the impaired penetration could be completely restored by adding soluble DEFB126 back to the sperm surface (19). Sperm surface charge appears to be critically important for cervical mucus penetration in the macaque (19), and much of the surface charge is contributed by the O-linked oligosaccharides that extend from the carboxyl half of the DEFB126 glycoprotein (16). Based on the macaque studies, we suggest that the change in composition of the human sperm glycocalyx associated with del/del genotype is due to a deficit of DEFB126, which results in loss of sperm surface properties important for penetration of negatively charged, viscous gel matrices.

Unexplained infertility is relatively common and often results in costly protracted clinical evaluations and emotional stress. The functional consequences reported here of the del/del variant of human *DEFB126* may provide new insight into factors contributing to male infertility. Given that the apparent proportion of men bearing the del/del *DEFB126* genotype in all populations we evaluated approaches 25%, it is not surprising that the genetic variant does not result in sterility. Rather, we propose that couples in which males possess the del/del genotype experience delays in achieving conception, because of reduced sperm performance in the environment of the cervix. This supposition is supported by independent data indicating that reduced sperm-cervical mucus penetration is significantly correlated with lower per cycle

conception rates (47, 48, 49, 50). As sperm must also pass through the mucin-rich utero-tubal junction and oviductal isthmus, and the highly visco-elastic HA matrix of the cumulus oophorus, the del/del genotype may also have implications for sperm function in the upper reproductive tract. In combination with female and or other male factors, sperm from males with a *DEFB126* del/del genotype could result in delays in achieving conception well beyond the one-year fertility benchmark.

Although our results would suggest that there is a strong selective pressure against the *DEFB126* variant allele, this allele is common in the Asian, European and African populations that we analyzed (Table 1). These observations suggest that the allele is old and has been maintained in the human population by balancing selection. For example, reproductive success and time to conception appear to be the same for males with wt/wt and del/wt genotypes; this may indicate that there is a selective advantage for heterozygotes as a result of another function of the del allele. In support of this, when the populations are analyzed as a group and using a sensitive test to detect a higher than expected heterozygote frequency, heterozygotes are indeed more frequent than expected given the allele frequencies and Hardy-Weinberg equilibrium ($p=0.0375 +/- 0.0003$). However, the 2-nucleotide deletion reported here generates a non-stop mRNA. The evidence that non-stop mRNAs yield null alleles (26, 27, 31) is consistent with the apparent lack of sperm surface glycosylation suggested by our lectin studies. One explanation for the high frequency of the del allele could be that lower expression of DEFB126 in the heterozygote confers some selective advantage, perhaps by changing the manner in which other epididymal proteins interact with the sperm surface. An alternative possibility is that a variant gene product is expressed from the del allele that cannot adhere to sperm, but that provides some

benefit to fecundity. Further studies will be required to determine how the del allele in *DEFB126* could provide the apparent selective advantage for heterozygotes.

In conclusion, our results point to a potential cause of impaired human fertility, and consequently the possibility of new clinical treatments. Genotype analysis of sub-fertile males for the *DEFB126* deletion polymorphism could assist in determining the most efficient steps for fertility interventions. By establishing genotype early in the infertility evaluation, clinicians could use this scientific evidence to justify rapid progression to directed interventions such as intrauterine insemination and *in vitro* fertilization, thus saving couples the time and expense of a protracted workup. In addition, on the basis of experiments in cynomolgus macaques, addition of a glycosylated recombinant DEFB126 to deficient sperm might augment other therapeutic approaches to infertility such as vaginal or cervical artificial insemination. Further research is needed to better understand the ramifications of this sequence variant in reproductive physiology.

Materials and Methods

qRT-PCR Analysis

Total RNA from epididymal tissue (lots A703139 and A703144) was obtained from Biochain Institute, Inc. For cDNA synthesis, 5 mg of total RNA was reverse transcribed with Superscript II reverse transcriptase (50 units) with an oligo- $(dT)_{12-18}$ primer as described (51). Real-time PCR was performed with the single-stranded epididymal cDNA and oligonucleotide primer pairs *DEFB125*-220s/404a, *DEFB126*-199s/330a, *DEFB129*-441s/546a and hGAPDH-597s/722a (Table 5) as described in Wehkamp et al. (51). Assays were performed in triplicate

(standard deviation between assays <10%). Values for β-defensins were normalized to expression of GAPDH mRNA as a control.

DEFB126 Genotype Analysis

DEFB126 genotype analysis on sperm was performed with isolated genomic DNA (10 ng) as a template in a standard PCR reaction using primer pairs oligonucleotide primer pairs DEFB126-154s/409a (Table 5). The DNA product was then subjected to dideoxysequence analysis using DEFB126-278s as a primer.

In the prospective cohort study, DNA was extracted from blood lymphocytes using standard procedures, and husbands were genotyped for the *DEFB126* sequence variant using single tube bi-directional allelic specific amplification. We designed wild-type (wt)-allele specific primers (Forward: 5'-AAGGGACTGCTGT GTTCCAG-3'; backward: 5'-ACCAGTGGGAGAAACGGGCGT-3') for amplification of 169 bp fragments from the homozygous wt/wt genotype. Similarly, we designed deletion (DEL)-allele specific primers (Forward: 5'-CTTCGATGGCTCCTACGCG-3'; Backward: 5'-GCTGTGGGC CTAGAACTGTC-3') for amplification of 295 bp fragments from the homozygous del/del genotype. We performed PCR amplification in a volume of 10 μl containing 60 ng of genomic DNA, 10X PCR reaction buffer, 1mM $MgCl_2$, 200 μM of each deoxynucleotide triphosphate, 200 nM of each primer, and 0.25U Taq DNA polymerase using the GeneAmp PCR system 2700 (Applied Biosystems). The PCR cycle included an initial denaturation at 94°C for 3 min, amplification for 38 cycles consisting of denaturation at 94°C for 1 min, annealing at 57°C for 1 min, and extension at 72°C for 1 min followed by a final extension step at 72°C of 7 min. We resolved PCR fragments using electrophoresis on 2% agarose gels and ethidium bromide staining. For quality control, we genotyped 10% of samples using separate tubes for wt-allele

and del-allele specific primers. Some genotype assignments were based on PCR amplification of a 250 bp segment of genomic DNA encompassing the dinucleotide deletion, followed by direct dideoxy sequence analysis of the PCR product. Other genotypes were assigned by using a modified primer to generate an *Mnl*I restriction enzyme site on the wt allele, followed by enzyme digestion. Genomic DNA (5-10 ng) was amplified in 10uL 1xKappa A PCR buffer (Kapa Biosystems), including a final concentration of 1.5mM $MgCl_2$, 0.2mM dNTP, 0.5µM of each primer (reverse 5'- GTTCAACCAGTGGGAGAAACGAG-3' with 5'-hydroxyl labeled with fluorescent dye HEX, forward 5'-AGAATGGTTGGGCAATGTGC-3'), 0.5units *Taq* DNA polymerase (Kapa Biosystems). Amplification conditions were 95°C for 30 seconds, 61°C for 30 seconds, 70°C for 30 seconds, for a total of 26 cycles. To the 10 µL of PCR product, 2µl of appropriate restriction enzyme buffer (1xNEB Buffer 2) containing 0.5u of *Mnl*I (New England Biolabs) was added and incubated at 37°C for at least 16 hours. An aliquot (1µL) of the digest was analyzed by capillary electrophoresis (ABI3130xl) according to the manufacturer's instructions, and fragment analysis using Genescan software.

Sperm preparation and analysis

Semen samples from 21 donors were collected by masturbation into non-spermatoxic specimen containers and allowed to liquefy for 20 minutes before use. A small fraction (<100 µl) of each sample was used for *DEFB126* genoptyping. Two of these donors moved out of the vicinity before semen evaluations could be performed. Semen samples from 19 of the donors were evaluated for volume, sperm density, percentage of motile sperm, and total numbers of motile sperm (Table 2) according to the WHO '87 methods (52). Semen from three of the donors had total motile sperm counts that were consistently too low for performance of HA penetration assays. For the 16 remaining donors, 10 µl drops of semen was layered onto glass

345 slides, dried, and stained using the method of Papanicolaou as described by Katz et al. (53) . Two hundred sperm per male per treatment were scored for various categories of abnormal forms (Table 2) according to the WHO '87 method (52). For computer assisted sperm analysis (CASA) and HA penetration experiments, sperm were washed twice by centrifugation (~300 x g) in modified BWW medium with 0.3% BSA (mBWW) and resuspended into mBWW at a motile
350 sperm concentration of 25 x 10^6/ml.

For measurements of sperm motion characteristics with CASA, videomicrography was performed as described (39, 54). Briefly, 4µl drops of sperm suspension were loaded into two µ-cell semen analysis chambers (Fertility Technologies Inc.) with a 10 micron depth. In each chamber, 8-10 randomly selected microscope fields were video recorded, capturing several
355 hundred sperm. Motion characteristics of the recorded sperm were analyzed using the HTM Ceros, version 10.9d (Hamilton Thorne Biosciences, Inc.). Sperm tracks were digitally captured using a frame rate of 60Hz and a minimum track time of 1 sec. At least 200 sperm per semen sample were analyzed for curvilinear velocity (VCL), straight-line velocity (VSL), and amplitude of lateral head displacement (ALH) (Table 2).

360 A 20 micron-deep slide chamber containing HA was prepared for HA penetration experiments as described previously for HA and cervical mucus penetration assays (19, 39). HA gel was composed of 5 mg purified hyaluronate (220 kDa fraction) per ml of HEPES-buffered BWW medium supplemented with 3% BSA. The slide chamber was warmed for 5 min on a microscope stage warmer (Motion Analysis, Inc.) set at 37^{o}C prior to the addition of sperm.
365 Twenty µL of sperm samples were introduced to the open side of the HA chamber and were immediately drawn by capillary action to the HA interface. Sperm were observed in HA with an Olympus BH2 microscope and a 10X phase objective and video recorded as described

previously (19, 39). After 2 min from the time sperm were introduced to the chamber, video recordings were initiated, capturing a region in the center of the microscope field that was approximately 2.75mm from the sperm-HA interface. Recordings continued for a minimum of 4 min. HA penetration was quantified from video recordings by counting the number of sperm in the video field that was paused at the very beginning of the recording (t = 2 min) and every min thereafter (t = 3-6 min) of the 4 min recording interval.

Measures of HA penetration, CASA, and sperm morphology were determined for three semen samples from each donor. Rates of HA penetration, VCL, and % normal forms were averaged for each donor and analyzed by genotype with 1-way ANOVA (α=0.05) followed by Tukey multiple range testing. All data met assumptions of normality of distribution and homogeneity of variance as determined with the Shapiro-Wilk test and Levene's test, respectively (55).

Lectin-labeling studies were performed on sperm from 15 of the donors (one of the remaining 16 donors left the program prior to initiation of the lectin experiments). Donor semen was washed over 40% Percoll to remove the majority of white cells and then washed by centrifugation in mBWW. Total sperm concentration was adjusted to 5 x 10^6 sperm/ml sperm and sperm were fixed with 1% paraformaldehyde/0.1% glutaraldehyde for 30 min. Sperm were washed repeatedly in DPBS, treated with neuraminidase (0.5 units/5 million sperm), washed into blocking solution and incubated with FITC- conjugated lectin ABA as described by Yudin et al. (16). Digital micrographs of sperm were digitally captured and fluorescence intensity analyzed using MetaMorph 6.1 Image Analysis (Universal Imaging Corp.) software as described by Tollner et al. (17, 45). Fluorescence (pixel) intensity data for individual sperm were averaged for each donor and analyzed by genotype. As the standard deviations of lectin labeling intensity

were roughly proportional to the means determined for each genotype, data initially did not meet assumptions of homogeneity of variance (Leven's test; $p = 0.024$). Following log transformation of pixel intensity data all assumptions of the ANOVA were met. Transformed data were analyzed by genotype with 1-way ANOVA ($\alpha = 0.01$). Differences between genotypes in mean average pixel intensity were further evaluated with Tukey multiple range testing. Analyses were conducted with SAS statistical program (SAS Institute), according to the principles described by Steel, Torrie, and Dickey (55).

Prospective cohort study

The protocols for the prospective cohort study were approved prior to implementation by the institutional review boards of the Harvard School of Public Health and the Anhui Medical University Institute of Biomedicine. Approval for secondary analysis of data from human subjects and preparation of this manuscript was obtained from Simon Fraser University.

After obtaining contact information from registrations of marriages with the provincial government and planned pregnancies with the family planning bureau, we contacted couples at their homes. After obtaining oral consent, we explained the study and invited eligible couples to participate. The inclusion criteria were 1) the marriage was the first for both the wife and husband; 2) the wife's age was between 20 and 34 years; 3) the wife was not a smoker and had never been one in the past; 4) both the wife and husband were available for the study; and 5) the couple currently lived together or planned to live together after marriage. Couples were eligible for inclusion in the study if they planned to stop contraception (or begin sexual activity) and try to conceive in the near future. Those who agreed to participate were invited to a field office at a later date for baseline procedures, which commenced only after the study was explained again in

detail, couples had an opportunity to receive answers to any questions they had about the study, and both signed a written consent.

This study was originally designed to investigate gene-environment interactions associated with pesticide exposure and human fertility and healthy pregnancy. We originally planned to study young couples who were farmers and therefore would have been exposed to pesticides. However, due to rapid economic changes, when our field operations began most young couples from this agricultural region had begun migrating to urban centers for non-agricultural employment and so did not have occupational exposures to pesticides. According to custom, most returned to their native homes during the period around the lunar New Year and it was common for young couples to register their marriages and marry during this period. As a consequence, 75% of our recruitment occurred within the 60 days prior to the lunar New Years which were January 22, 2004 and February 9, 2005. All subjects who met the inclusion criteria were invited to participate in the study and more than 90% took part and completed the study.

We attempted to contact all participants in their urban homes by telephone in November, 2006, which was 21 months after the last couples were enrolled. Our rationale for this follow-up period was to allow up to 12 months for couples to achieve pregnancy (the clinical cutoff for a definition of infertility) plus 9 months to allow follow-up of the birth outcome. Out of 812 enrolled couples, we successfully contacted 749 (92%). All participants had either achieved pregnancy or attempted conception for at least 21 months prior to when we contacted them.

In the 749 couples with follow-up data, the mean (standard deviation) age and body-mass index among men were 25.9 (2.6) and 21.6 (2.6) and among women were 23.4 (2.3) and 21.2 (2.6), respectively. The prevalence of smoking was 56% among men and 1% among women (however, these 4 women who reported smoking were excluded from our analysis). Considering

these 749 as the total group of participants, we used ANOVA models to test for differences in the mean values of age and body-mass index (both of which were approximately normally distributed) between couples included and excluded in our models of the odds or self-reported pregnancy (included n=509) and time to live birth (included n=480). (See Results section of main text for details about exclusions.) At α=0.10, there were no statistically significant differences between included and excluded men and women in either model for mean age or body-mass index. Nor were there differences between included and excluded men in either model for the prevalence of active smoking by chi-square test.

We used logistic regression to model the relative odds of self-reported pregnancy at follow-up by male *DEFB126* genotype in 509 couples, which is appropriate for our binary outcome and categorical predictor (56). The odds of self-reported pregnancy within a genotype group are calculated as the number who achieved pregnancy divided by the number who did not. We report the odds ratio (and its two-sided 95% confidence interval) of self-reported pregnancy, which is calculated as the odds of self-reported pregnancy in one genotype group (for example del/del) divided by the odds in another group (for example wt/wt and wt/del combined).

We also used Cox proportional hazards regression to model the relative hazard of live birth at follow-up by male *DEFB126* genotype in 480 couples (after excluding 29 with spontaneous or induced abortion). We did not have accurate information about the timing of the last menstrual period prior to pregnancy so could not model the relative hazard of pregnancy. The hazard was defined as the probability of live birth occurring per month of follow-up. We report the hazard ratio (and its two-sided 95% confidence interval) of live birth, which is calculated as the hazard of live birth in one genotype group (for example del/del) divided by the hazard in another group (for example wt/wt and wt/del combined). We used a plot of the

standardized score process against time to live birth (57) to confirm that the ratio of the hazard functions between genotype groups was constant over follow-up time, which is an assumption of this model.

Testing for departure from Hardy-Weinberg Equilibrium

In the Chinese fertility cohort, comparison of genotype frequencies with those given Hardy Weinberg expectations was performed using the Pearson goodness-of-fit test (chi-squared test), primarily as a check of genotyping accuracy. This test is rather conservative and relies on rejection of the null hypothesis. To detect any excess of heterozygotes on a larger cohort of individuals, we used the population genetics software GENEPOP v. 4 (58). This specifically tests for a higher than expected heterozygote count across all populations, without pooling populations, with standard error estimated by Markov-chain Monte-Carlo methods (58). This "exact test" is more powerful than goodness-of-fit tests, and because rather than just rejecting the null hypothesis, we specify an alternative hypothesis, i.e. a higher than expected frequency of heterozygotes, the U-test can be used (59)

References

1. P. J. Rowe, F. H. Comhaire, T. B. Hargreave, H. J. Mellows, WHO Manual for the Standardized Investigation and Diagnosis of the Infertile Couple. (Cambridge University Press, Cambridge, 1993).
2. P. J. Rowe, F. H. Comhaire, T. B. Hargreave, A. M. Mahmoud, WHO Manual for the Standardized Investigation, Diagnosis and Management of the Infertile Male. (Cambridge University Press, Cambridge, 2000).
3. R. Strickler, Factors Influencing Fertility. In *Fertility Evaluation and Treatment.* W. Keye, R. Chang, R. Rebar, M. Soules, Eds. 8-18. (W.B. Saunders, Philadelphia, 1995)
4. D. S. Guzick, J. W. Overstreet, P. Factor-Litvak, C. K. Brazil, S. T. Nakajima, C. Coutifaris, S. A. Carson, P. Cisneros, M. P. Steinkampf, J. A. Hill, D. Xu, D. L. Vogel; National

Cooperative Reproductive Medicine Network, Sperm morphology, motility, and concentration in fertile and infertile men. *N. Engl. J. Med.* **345**, 1388-1393 (2001).

5. P. Jedrzejczak, G. Taszarek-Hauke, J. Hauke, L. Pawelczyk, A. J. Duleba, Prediction of spontaneous conception based on semen Parameters. *Int. J. Androl*. **31**, 499–507 (2005).

6. A. Agarwal, R. K. Sharma, D. R. Nelson, New Semen Quality Scores Developed by Principal Component Analysis of Semen Characteristics. *J. Androl.* **24**, 343-352 (2003).

7. J. Collins, Unexplained Infertility. In *Infertility Evaluation and Treatment*. W. Keye, R. Chang, R. Rebar, M. Soules, Eds. 249-262 (W.B. Saunders, Philadelphia, 1995).

8. S. Schroter, C. Osterhoff, W. McArdle, R. Ivell, The glycocalyx of the sperm surface. *Hum. Reprod. Update* **5**, 302-313 (1999).

9. A. B. Diekman, Glycoconjugates in sperm function and gamete interactions: how much sugar does it take to sweet-talk the egg? *Cell. Mol. Life. Sci*. **60**, 298-308 (2003).

10. J. Pela´ez, J. A. Long, Characterizing the Glycocalyx of Poultry Spermatozoa: II. In Vitro Storage of Turkey Semen and Mobility Phenotype Affects the Carbohydrate Component of Sperm Membrane Glycoconjugates. *J. Androl.* **29**, 431-439 (2008).

11. I. Jiménez, H. Gonzalez-Marquez, R. Ortiz, M. Betancourt, J. Herrera, R. Fierro, Expression of lectin receptors on the membrane surface of sperm of fertile and subfertile boars by flow cytometry. *Arch. Androl.* **48**, 159-166 (2002).

12. S. Purohit, M. Laloraya, P. G. Kumar, Distribution of N- and O-linked oligosaccharides on surface of spermatozoa from normal and infertile subjects *Andrologia* **40**, 7–12 (2007).

13. L. K. Gabriel, D. R. Franken, Binding of human spermatozoa to lectin-coated agarose microbeads. *Arch Androl*. **38**, 133-41 (1997).

14. R. I. Lehrer, Primate defensins. *Nat Rev Microbiol.* **2**, 727-738 (2004).

15. A. I. Yudin, T. L. Tollner, M. W. Li, C. A. Treece, J. W. Overstreet, G. N. Cherr, ESP13.2, a member of the beta-defensin family, is a macaque sperm surface-coating protein involved in the capacitation process. *Biol. Reprod.* **69,** 1118-1128 (2003).

16. A. I. Yudin, C. A. Treece, T. L. Tollner, J.W. Overstreet, G. N. Cherr, The Carbohydrate element of ESP13.2/DEFB126, the major component of the cynomolgus macaque sperm plasma membrane glycocalyx. *J Membr Biol.* **207**, 119-129 (2005).

17. T. L. Tollner, A. I. Yudin, A. F. Tarantal, C. A. Treece, J. W. Overstreet, G. N. Cherr, Beta-defensin 126 on the surface of macaque sperm mediates attachment of sperm to oviductal epithelia. *Biol. Reprod.* **78,** 400-412 (2008).

18. A. I. Yudin, S. E. Generao, T. L. Tollner, C. A. Treece, J. W. Overstreet, G. N. Cherr, Beta-defensin 126 on the cell surface protects sperm from immunorecognition and binding of anti-sperm antibodies. *Biol. Reprod.*, **73,** 1243-1252 (2005).

19. T. L. Tollner, A. I. Yudin, C. A. Treece, J. W. Overstreet, G. N. Cherr, Macaque sperm coating protein DEFB126 facilitates sperm penetration of cervical mucus. *Hum. Reprod.* **23**, 2523-2534 (2008).

20. A. Yudin, T. Tollner, C. Treece, R. Kays, G. Cherr, J. Overstreet, C. Bevins, Beta-defensin 22 is a major component of the mouse sperm glyocalyx. *Reprod.* **136**, 753-765 (2008).

21. F. W. Avila, L. K. Sirot, B. A. LaFlamme, C. D. Rubinstein, M. F. Wolfner, Insect Seminal Fluid Proteins: Identification and Function. *Annu Rev Entomol*. **56,** 21-40 (2011)

22. S. Okuda, H. Tsutsui, K. Shiina, S. Sprunck, H. Takeuchi, R. Yui, R. D. Kasahara, Y. Hamamura, A. Mizukami, D. Susaki, N. Kawano, T. Sakakibara, S. Namiki, K. Itoh, K. Otsuka, M. Matsuzaki, H. Nozaki, T. Kuroiwa, A. Nakano, M. M. Kanaoka, T. Dresselhaus, N. Sasaki, T. Higashiyama, Defensin-like polypeptide LUREs are pollen tube attractants secreted from synergid cells. *Nature*. **458**, 357-361 (2009).

23. S. Amien, I. Kliwer, M. L. Márton, T. Debener, D. Geiger, D. Becker, T. Dresselhaus, Defensin-like ZmES4 mediates pollen tube burst in maize via opening of the potassium channel KZM1. *PLoS Biol.* **8**, e1000388 (2010).

24. E. J. Hollox, J. A. Armour, Directional and balancing selection in human beta-defensins. *B.M.C. Evol. Biol*. **8**, 113 (2008).

25. Y. Radhakrishnan, K. G. Hamil, S. Yenugu, S. L. Young, F. S. French, S. H. Hall, Identification, characterization, and evolution of a primate beta-defensin gene cluster. *Genes Immun.* **6**, 203-210 (2005).

26. A. Chatr-Aryamontri, M. Angelini, E. Garelli, G. Tchernia, U. Ramenghi, I. Dianzani, F. Loreni, Nonsense-mediated and nonstop decay of ribosomal protein S19 mRNA in Diamond-Blackfan anemia. *Hum. Mutat.* **24**, 526-533 (2004).

27. A. Ameri, D. K. Machiah, T. T. Tran, C. Channell, V. Crenshaw, K. Fernstrom, M. Khachidze, A. Duncan, S. Fuchs, T. E. Howard, Nonstop mutation in the factor (F)X gene of a severely haemorrhagic patient with complete absence of coagulation FX. *Thromb. Haemost.* **98**, 1165-1169 (2007).

28. Frischmeyer, P.A., van Hoof, A., O'Donnell, K., Guerrerio, A.L., Parker, R. & Dietz, H.C. An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science*. **295**, 2258-2261 (2002).

29. L. E. Maquat, Molecular biology. Skiing toward nonstop mRNA decay. *Science*. **295**, 2221-2222 (2002).

30. O. Isken, L. E. Maquat, Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes Dev* .**21**, 1833-1856 (2007).

31. N. Akimitsu, J. Tanaka, J. Pelletier, Translation of nonSTOP mRNA is repressed post-initiation in mammalian cells. *EMBO J.* **26**, 2327-2338 (2007).

32. R. J. Aitken, Sperm function tests and fertility. *Int. J. Androl.* **29**, 69-75 (2006)

33. I. K. Gipson, S. B. Ho, S. J. Spurr-Michaud, A. S. Tisdale, Q. Zhan, E. Torlakovic, J. Pudney, D. J. Anderson, N. W. Toribara, J. A. Hill, Mucin genes expressed by human female reproductive tract epithelia. *Biol. Reprod.* **56**, 999-1011 (1997).

34. E. Lagow, M. M. DeSouza, D. D. Carson, Mammalian reproductive tract mucins. *Hum. Reprod. Update* **5**, 280-292 (1999).

35. I. Gatej, M. Popa, M. Rinaudo, Role of the pH on Hyaluronan Behavior in Aqueous Solution. *Biomacromol.* **6**, 61-67 (2005).

36. S. Tang, C. Garrett, H. W. Bake, Comparison of human cervical mucus and artificial sperm penetration media. *Hum. Reprod.* **14**, 2812-2817 (1999).

37. J. Neuwinger, T. G. Cooper, U. A. Knuth, E. Nieschlag, Hyaluronic acid as a medium for human sperm migration tests. *Hum. Reprod.* **6**, 396-400 (1991).

38. R. J. Aitken, H. Bowie, D. Buckingham, D. Harkiss, D. W. Richardson, K. M. West, Sperm penetration into a hyaluronic acid polymer as a means of monitoring functional competence. *J. Androl.* **13**, 44-54 (1992).

39. T. L. Tollner, Q. Dong, C. A. VandeVoort, Frozen-thawed rhesus sperm retain normal morphology and highly progressive motility but exhibit sharply reduced efficiency in penetrating cervical mucus and hyualuronic acid gel. *Cryobiology* **62**, 15-21 (2011)

40. R. J. Aitken, M. Sutton, P. Warner, D. W. Richardson, Relationship between the movement characteristics of human spermatozoa and their ability to penetrate cervical mucus and zona-free hamster oocytes. *J. Reprod. Fertil*. **73,** 441-449 (1985).

41. D. Mortimer, I. J. Pandya, R. S. Sawers, Relationship between human sperm motility characteristics and sperm penetration into human cervical mucus in vitro. *J. Reprod. Fertil.* **78,** 93-102 **(**1986).

42. W. Eggert-Kruse, J. Reimann-Andersen, G. Rohr, S. Pohl, W. Tilgen, B. Runnebaum, Clinical relevance of sperm morphology assessment using strict criteria and relationship with sperm-mucus interaction in vivo and in vitro. *Fertil. Steril.* **63**, 612-624 (1995).

43. P. Morales, D. F. Katz, J. W. Overstreet, S. J. Samuels, R. J. Chang, The relationship between the motility and morphology of spermatozoa in human semen. *J. Androl.* **9,** 241-247 **(**1988).

44. D. J. Balding, A tutorial on statistical methods for population association studies. *Nat. Rev. Genet.* **7**, 781-791 (2006).

45. T. L. Tollner, A. I. Yudin, C. A. Treece, J. W. Overstreet, G. N. Cherr, Macaque sperm release ESP13.2 and PSP94 during capacitation: The absence of ESP13.2 is linked to sperm-zona recognition and binding. *Mol. Reprod. Dev.* **69,** 325-337 (2004).

46. T. Tollner, , C. VandeVoort, A. Yudin, C. Treece, J. Overstreet, G. Cherr, Release of DEFB126 from macaque sperm and completion of capacitation are triggered by conditions that simulate periovulatory oviductal fluid. *Molec. Reprod. Dev.* **76,** 431-443 (2009).

47. W. Eggert-Kruse, G. Leinhos, I. Gerhard, W. Tilgen, B. Runnebaum, Prognostic value of in vitro sperm penetration into hormonally standardized human cervical mucus. *Fertil. Steril.* **51**, 317-323 (1989).

48. A. N. Beltsos, S. Fisher, M. L. Uhler, E. D. Clegg, M. Zinaman, The relationship of the postcoital test and semen characteristics to pregnancy rates in 200 presumed fertile couples. *In.t J. Fertil. Menopausal. Stud.* **41**, 405-411 (1996).

49. C. M. Glazener, W.C. Ford, M. G. Hull, The prognostic power of the post-coital test for natural conception depends on duration of infertility. *Hum. Reprod.* **15**, 1953-1957 (2000).

50. C. C. Hunault, J. S. Laven, I. A. van Rooij, M. J. Eijkemans, E. R. te Velde, J. D. Habbema, Prospective validation of two models predicting pregnancy leading to live birth among untreated subfertile couples. *Hum. Reprod.* **20**, 1636-1641 (2005).

51. J. Wehkamp, H. Chu, B. Shen, R. W. Feathers, R. J. Kays, S. K. Lee, C. L. Bevins, Paneth cell antimicrobial peptides: topographical distribution and quantification in human gastrointestinal tissues. *FEBS Lett.* 580, 5344-5350 (2006).

52. World Health Organization. WHO laboratory manual: the examination of human semen and sperm-cervical mucus interaction. 2nd Edition. (Cambridge University Press, Cambridge, 1987).

53. D. F. Katz, J. W. Overstreet, S. J. Samuels, P. W. Niswander, T. D. Bloom, E. L. Lewis, Morphometric analysis of spermatozoa in the assessment of human male fertility. *J. Androl.* **7,** 203-210 (1986)

54. C. A. VandeVoort, T. L. Tollner, J. W. Overstreet, Separate effects of caffeine and dbcAMP on macaque sperm motility and interaction with the zona pellucida. *Mol. Reprod. Dev.* **37**, 299-304 (1994).

55. R. G. Steel, J. H. Torrie, D. A. Dickey, Principles and Procedures of Statistics, a Biometrical Approach. New York: WCB/McGraw-Hill (1997).

56. V. Bewick, L. Cheek, J. Ball, Statistics review 14: Logistic regression. *Crit. Care* **9**, 112-118 (2005).

57. D. Y. Lin, L. J. Wei, Z. Ying, Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**, 557–572 (1993).

58. S. W. Guo, E. A. Thompson, Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* **48**, 361-372 (1992).

59. F. Rousset, M. Raymond, Testing heterozygote excess and deficiency. *Genetics* **140**, 1413-1419 (1995)

60. Acknowledgements: We thank Kerstien Padgett for technical assistance.Funding: This work was supported, in part, by grants from the National Institutes of Health (R01ES008957, S.A.V. and X.X., K01ES012052, S.A.V., and R01AI32738, C.L.B.), and from the National Science Foundation (IOS-0843649, G.N.C.). E.J.H is supported by a Medical Research Council New Investigator award GO801123 and Wellcome Trust grant 087663. **Author Contributions:** T.L.T., A.I.Y., J.W.O., G.N.C. and C.L.B. conceived the study. T.L. and R.J.K. performed cloning, genotyping and sequence analysis. T.L.T. managed the semen donor program, performed semen evaluations and sperm bioassays, and analyzed data on the effect of the *DEFB126* polymorphism on sperm form and function. E.J.H. performed population genotyping and data analysis. X.L., G.T., H.X., and X.X. planned and supervised all of the prospective cohort study. S.A.V. supervised field work, conducted all data analysis and wrote the manuscript related to the prospective cohort study. S.A.V., T.L.T., E.J.H., C.L.B. and G.N.C. wrote the manuscript. All authors reviewed and commented on the final version of the manuscript. **Competing interests:** E.J.H, R.J.K., X.L., G.T., and X.H., declare no competing financial interests. T.L.T., S.A.V., A.I.U., T.L., J.W.O., X.X., G.N.C. and C.L.B. are co-

inventors named in a provisional patent related to the subject of this manuscript. **Accession numbers:** Nucleotide sequences are deposited in the NCBI at AK22598 and rs11468374.

Figure Legends

Figure 1. Deduced amino acid sequence alignment of DEFB126 from four primate species. The deduced amino acid sequence for human (*Homo sapiens*, accession number NP_112193), common chimpanzee (*Pan troglodytes*, XP_514453), Gorilla (*Gorilla gorilla*, A4H243.1) and Lar Gibbon (*Hylobates lar*), A4H245.1) are aligned, with identity to human sequence "-" and gap for maximum alignment "_" noted.

Figure 2. Nucleotide and deduced amino acid sequence of a common human *DEFB126* sequence variant. **A**. Schematic of *DEFB126* mRNA showing site and consequence of common 2-nucleotide deletion variant. The 2-nucleotide deletion predicts a frame shift and a variant reading frame lacking an in-frame stop codon. **B**. Representative dideoxysequence sequence analysis chromatogram of wildtype and deletion variant alleles. Shown is *DEFB126* sequence analysis of human epididymal cDNA clones. The *DEFB126* variant had a 2-nucleotide omission (deletion), causing a frame-shift in the open reading frame of *DEFB126*. **C**. Quantitative RT-PCR analysis of epididymal mRNA from an individual with wt/wt (open bar) and an individual with del/del (solid bar) genotypes. Assays specific for mRNA of three β-defensins expressed in the epididymis (*DEFB125*, *DEFB126*, and *DEFB129*) were normalized to expression of *GAPDH* mRNA as a control. The epididymal specimen with the del/del genotype has reduced levels (approximately 10-fold) of *DEFB126* expression, consistent with published analysis of mutations in other genes whose mRNA lack in-frame stop codons (26, 27, 28), where such aberrant mRNA

are less abundant than corresponding wild-type mRNA.

Figure 3: Sperm surface O-linked oligosaccharides as determined by labeling with ABA lectin. Sperm from donors with *DEFB126* del/del genotype exhibit reduced surface Sperm O-linked oligosaccharides. Human sperm were treated with neuraminidase, fixed, and incubated with FITC- conjugated lectin ABA (see methods). Micrographs of sperm fluorescence (right panel) and corresponding phase contrast (left panel) are shown for each donor. Sperm from del/del donors (rows 1-4) exhibit lower ABA label intensity than sperm from wt/del donors (rows 5-8) and wt/wt donors (rows 9-12). Scale bar (left panel of row 1) equals 10 microns.

Figure 4: *DEFB126* genotype and sperm penetration of HA gel. Sperm from donors with del/del genotype exhibit reduction in HA penetration that is consistent with ABA-lectin intensity but not with sperm morphology or progressive motility. Sperm from donors genotyped for the *DEFB126* polymorphism were used in HA penetration experiments. A. HA penetration: sperm penetration of HA gels was measured as average number of sperm penetrating 2.75mm past the sperm-gel interface at 1 min intervals. (B) Curvilinear velocity: sperm suspensions were analyzed by CASA for average curvilinear velocity (VCL). (C) Morphology: sperm morphology was determined according to WHO ‘87 method (52) and reported as average percent normal forms (% normal). Observations reported in A-C were paired and represent data averaged across three ejaculates (sub-samples) from each donor. (D) ABA lectin labeling intensity was determined with Metamorph software for sperm from genotyped donors. Data reported as means+/-sem. Crosses (+) ,asterisks (*), and t-strokes (Ŧ) indicate significant differences at $p = 0.030$, $p = 0.008$, and $p.= 0.0006$, respectively.

Figure 5. Sperm surface ABA lectin labeling intensity as a function of HA Penetration (HAP). Simple regression analysis ($r = 0.800$; $p = 0.003$) was performed by plotting the fluorescence intensity of ABA-labeled sperm (average pixel intensity) against the rate of sperm HAP (sperm/min) by donor. Dark blue, light blue, and red plots indicate average values for sperm from wt/wt, wt/del, and del/del donors, respectively.

Tables

Table 1. Genotype frequencies of *DEFB126* DEL variant in different populations.

Population Cohorts	n	wt/wt	wt/del	del/del
Chinese (fertility cohort)	638	0.29	0.51	0.19
Chinese (HapMap)	45	0.33	0.49	0.18
Japanese (HapMap)	45	0.22	0.56	0.22
Utah, of European origin (HapMap)	60	0.16	0.58	0.25
Yoruba from Nigeria (HapMap)	60	0.15	0.48	0.37
British	91	0.21	0.60	0.19

Table 2.

Common assessments of sperm from donors genotyped for DEFB126 polymorphism.

General Semen Parameters

	Volume	[Sperm]	% Motility	Tot. Motile Cnt.
del/del (n=6)	2.9 ± .4*	40.4 ± 5	50 .2 ± 5	62 ± 17
wt /del (n=9)	2.7 ± .3	78.7 ± 25	43.8 ± 6	77.3 ± 22
wt / wt (n=4)	3.1 ± .3	38.7 ± 13	52.6 ± 7	58 ± 17

Sperm Morphology

		% Abnormal Forms		
	% Normal	Heads	Tails	Other
del/del (n=6)	50.1 ± 4*	28.6 ± 2	19.2 ± 3	2.4 ± .7
wt /del (n=6)	55.2 ± 7	26.3 ± 4	15.8 ± 3	2.1 ± .5
wt / wt (n=4)	48.7 ± 6	30.9 ± 4	16.6 ± 3	4.1 ± .6

CASA Motion Parameters

	VCL	VSL	ALH	LIN
del/del (n=6)	76.2 ± 2*	34 ± 3	4.2 ± .2	2.4 ± .7
wt /del (n=6)	83.6 ± 4	39.4 ± 2	4.4 ± .2	2.1 ± .5
wt / wt (n=4)	79 ± 4	33.9 ± 3	4.7 ± .2	4.1 ± .6

***All table values = mean ± sem**

Table 3. Relative odds (OR) of pregnancy by husband's *DEFB126* genotype in prospective cohort study.

Husband *DEFB126*	n	Pregnancies n (%)	OR (95% CI)	2-sided *P*
Additive model				
WT/WT	156	128 (82%)	Referent	
WT/Del	251	200 (80%)	0.9 (0.5, 1.4)	.577
Del/Del	102	72 (71%)	0.5 (0.3, 0.9)	.032
Recessive model				
WT/WT and WT/Del	407	328 (81%)	Referent	
Del / Del	102	72 (71%)	0.6 (0.4, 0.9)	.029

Table 4. Relative probability (HR, hazard ratio) of birth per month by husband's *DEFB126* genotype in prospective cohort study.

Husband *DEFB126*	n	Live birth n (%)	Mean (SD) time to live birth or end of follow-up in months	HR (95% CI)	2-sided P
Additive model					
WT/WT	145	106 (73%)	15.4 (7.4)	Referent	
WT/Del	235	165 (70%)	15.8 (7.4)	0.9 (0.7, 1.2)	0.581
Del/Del	100	60 (60%)	17.4 (7.4)	0.7 (0.5, 1.0)	0.026
Recessive model					
WT/WT and WT/Del	380	271 (71%)	15.7 (7.3)	Referent	
Del / Del	100	60 (60%)	17.4 (7.4)	0.7 (0.5, 1.0)	0.026

Table 5. Oligonucleotides for PCR analysis

Name	Sequence
DEFB125-220s	5’- CGA CGA CCA GCA TTT CCT GTG ATT C -3’
DEFB125-404a	GGT GGC ATA GTA GTC TCG GGA GTA GTG G
DEFB126-154s	AAG AAT GGT TGG GCA ATG TGC
DEFB126-199s	GCA AAC AAA GGG ACT GCT GTG TTC C
DEFB126-278s	CAG CAA CAA CAA CTT TGA TGA TGA C
DEFB126-330a	AGG AGC CAT CGA AGA CAT CGA AGC
DEFB126-409a	CCA CAA TGC TTT AAT GAG TCG GG
DEFB129-441s	CCA TCA GCA CTA TGA CCC CAG GAC
DEFB129-546a	GTT GGC AGT ATG TTT GGT GGA GGT G
hGAPDH-597s	TGC CAT CAC TGC CAC CCA GAA G
hGAPDH-722a	ATG ACC TTG CCC ACA GCC TTG G

Figure 1

Human	**MKSLLFTLAVFMLLAQLVSGNWYVKKCLNDVGICKKKCKP (1-40)**
Common Chimpanzee	- -
Western Gorilla	- -
Lar Gibbon	- -
Human	**EEMHVKNGWAMCGKQRDCCVPADRRANYPVFCVQTKTTRI (41-80)**
Common Chimpanzee	**G** - - - - - - - - - -**T**- -**T**
Western Gorilla	- -
Lar Gibbon	- - **L** -**K**- - - - -**A** - - - - - - - -**T**
Human	**STVTATTAT_TTLMMTTASMSSMAPTPVSPTG (81-112)**
Common Chimpanzee	- - - - -**R**____- - - - **V** - - - - - - - - - - - - - - - - -
Western Gorilla	- - - - - - - - _ - - - - - - - - - - - **L** - - - - - - - - - -
Lar Gibbon	- - - - - -**A** - -**T** - - - **V**- - - - - - - - - - - - - - - - - -**S**

Figure 2

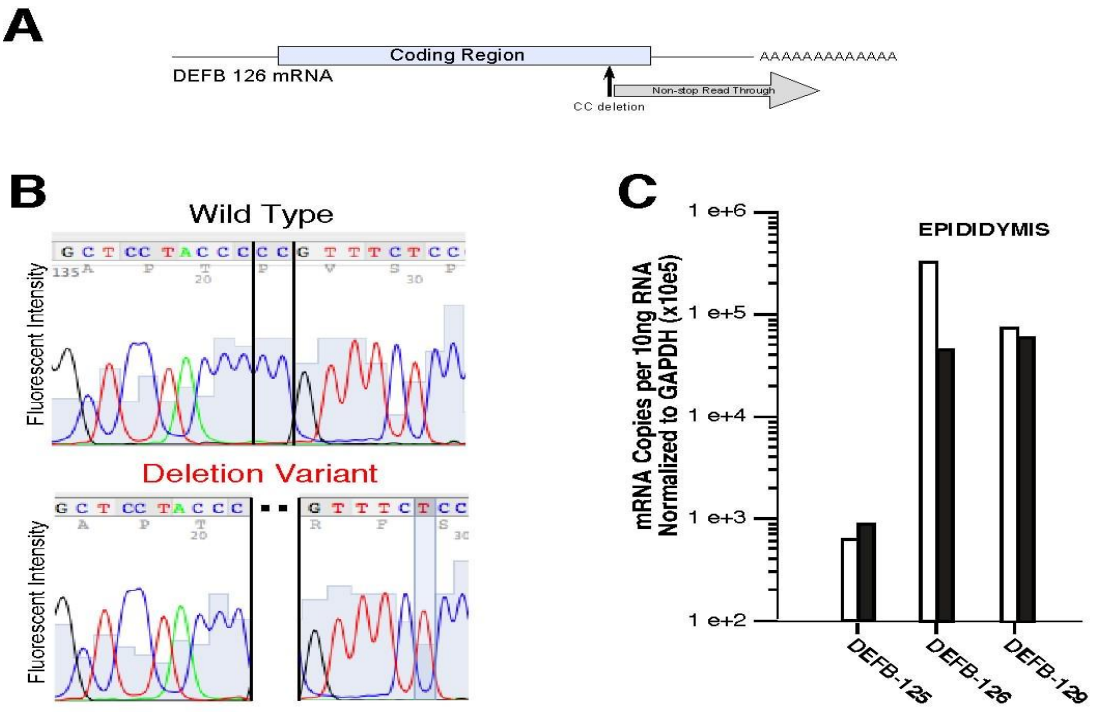


Figure 3

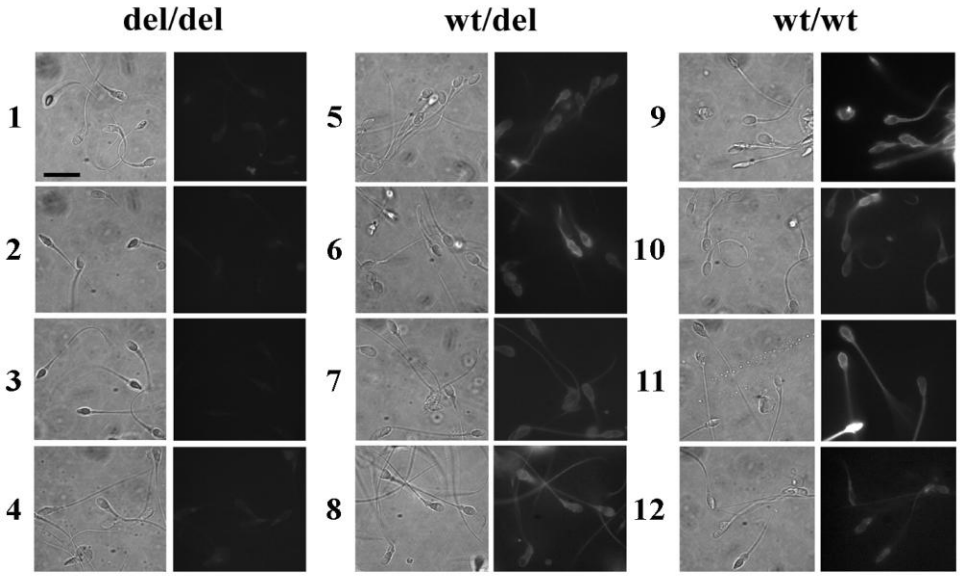
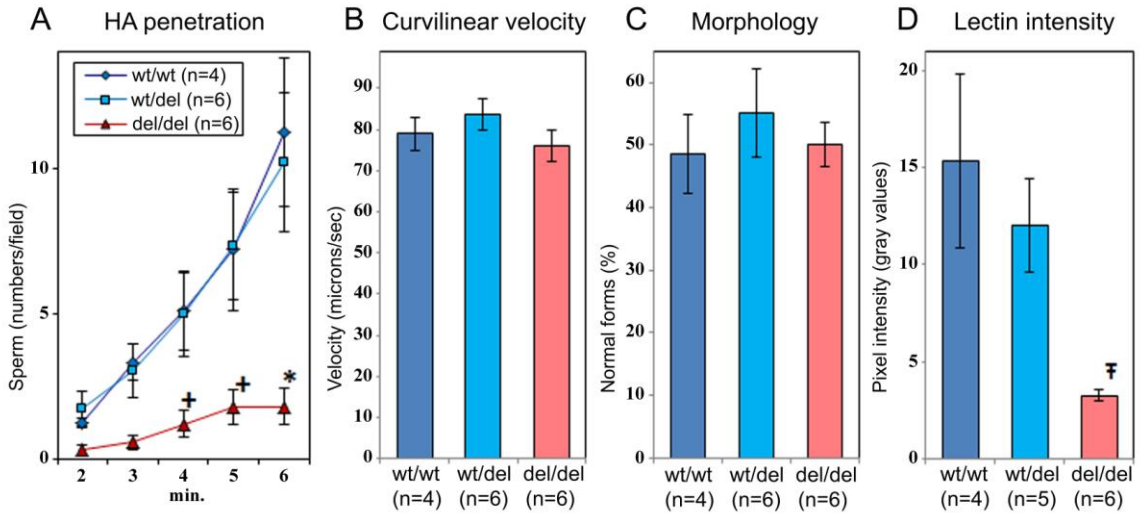


Figure 4



785

Figure 5