Identification of TEX101 functional interactome through proteomic measurement of human spermatozoa homozygous for the missense variant rs35033974

Christina Schiza^{1,2}, Dimitrios Korbakis^{1,3}, Keith Jarvi^{3,4}, Eleftherios P. Diamandis^{1,2,3,5*} and Andrei P. Drabovich^{1,2,5*}

¹Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Canada

²Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Canada

³Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Canada

⁴Department of Surgery, Division of Urology, Mount Sinai Hospital, University of Toronto, Toronto, Canada

⁵Department of Clinical Biochemistry, University Health Network, Toronto, Canada

Correspondence should be addressed to:

A. P. Drabovich, Ph.D., Department of Laboratory Medicine and Pathobiology, University of Toronto, 60 Murray Street; Room L6-201-4, Toronto, ON, M5T 3L9, Canada. Tel: (416) 586-4800 ext. 8805; e-mail: andrei.drabovich@utoronto.ca

or

E. P. Diamandis, Ph.D., M.D., Mount Sinai Hospital, 60 Murray St [Box 32]; Flr 6 - Rm L6-201-1, Toronto, ON, M5T 3L9, Canada. Tel: 416-586-8443; Fax: 416-619-5521; e-mail: eleftherios.diamandis@sinaihealthsystem.ca

Non-standard abbreviations: TEX101, Testis-expressed sequence 101 protein; LY6K, Lymphocyte antigen 6 complex, locus K; ADAM29, A disintegrin and metalloproteinase domain-containing protein 29; DPEP3, Dipeptidase 3; BH-adjusted t-test, Benjamini-Hochberg-adjusted t-test; FDR, False discovery rate; FWHM, Full width at half maximum; GPI, Glycosylphosphatidylinositol; LC-MS/MS, liquid chromatography - tandem mass spectrometry; LFQ, Label-free quantification; MS, Mass spectrometry; mAb, Monoclonal antibody; MWU, Mann Whitney Unpaired t-test; PRM, Parallel reaction monitoring; ROC AUC, Receiver operating characteristic area under the curve; SCX, strong cation exchange chromatography; SP, seminal plasma; SNV, Single nucleotide variation; SRM, Selected reaction monitoring; WT, wild-type.

Key words: TEX101; testis-expressed sequence 101 protein; LY6K; Lymphocyte antigen 6 complex, locus K; ADAM29, A disintegrin and metalloproteinase domain-containing protein 29; interactome; testis-specific proteins; spermatozoa; single nucleotide variant, unexplained male infertility; proteogenomics.

SUMMARY

TEX101 is a testis-specific cell-surface protein expressed exclusively in the male germ cells and a validated biomarker of male infertility. Mouse TEX101 was found essential for male fertility, and was suggested to function as a cell surface chaperone involved in maturation of proteins required for sperm migration and sperm-oocyte interaction. However, the precise functional role of human TEX101 is not known and cannot be studied *in vitro* due to the lack of human germ cell lines. Here, we genotyped 386 healthy fertile men and sub-fertile patients for a common and potentially deleterious missense variant rs35033974 of *TEX101*, and identified 52 heterozygous and 4 homozygous patients. We then discovered by targeted proteomics that the variant allele rs35033974 was associated with near-complete degradation (>97%) of the corresponding G99V TEX101 form, and suggested that spermatozoa of homozygous patients could serve as a knockdown model to study TEX101 function in humans. Differential proteomic profiling with label-free quantification measured 8,046 proteins in spermatozoa of eight men and identified 8 cell-surface and 9 secreted testis-specific proteins significantly down-regulated in four patients homozygous for rs35033974. Substantially reduced levels of testis-specific cell-surface proteins potentially involved in sperm migration and sperm-oocyte fusion (including LY6K and ADAM29) were confirmed by targeted proteomics and western blotting assays. Since recent population-scale genomic data revealed homozygous fathers with biological children, rs35033974 is not a single pathogenic factor of male infertility in humans. However, median TEX101 levels in seminal plasma were found 5-fold lower ($P=0.0005$) in heterozygous than in wild-type men of European ancestry. We conclude that spermatozoa of rs35033974 homozygous men have substantially reduced levels of TEX101 and could be used as a model to elucidate the precise TEX101 function, which will advance biology of human reproduction.

INTRODUCTION

Recent -omics studies identified 1,079 human genes with exclusive expression in testis (1). While function of many of testis-specific proteins is not known, it may be assumed that these proteins have unique and specialized roles in spermatogenesis and fertilization. Mutations, natural knockouts or deleterious single nucleotide variations in testis-specific genes could lead to severe male infertility with spermatogenesis arrest, low sperm concentration, abnormal spermatozoa morphology or low motility. In addition, mutations in testis-specific genes involved in sperm transit, acrosome reaction and sperm-oocyte interaction may result in no apparent phenotypical issues and male infertility which cannot be explained with routine diagnostics (2-4).

We previously discovered and validated a testis-specific protein TEX101 as a seminal plasma biomarker for the differential diagnosis of azoospermia and male infertility (5-9). The functional role of TEX101 is not known, but based on mouse models it was suggested as a spermatozoa cell-surface chaperone involved in the maturation of four cell-surface proteins from the ADAM family (10-12). *Tex101* knockout in mice resulted in absolute male sterility but normal sperm concentration, morphology and other phenotypical characteristics (10, 11). In the absence of TEX101 protein, ADAM 3-6 proteins were not properly processed and degraded. However, mouse data could not be translated into human studies since ADAM3, ADAM5 and ADAM6 genes are non-coding pseudogenes, while ADAM4 is not present in the human genome (13). Due to the lack of stable human male germ cell lines, TEX101-dependent proteins could not be identified in humans.

As an alternative, we suggested that the functional role of TEX101 could be studied in human clinical samples, such as spermatozoa. Our previous work on TEX101 levels in seminal plasma revealed a small population of men with high sperm count, but unexplained infertility and very low levels of TEX101 protein in seminal plasma and spermatozoa (8). In this work, we hypothesized that some genomic alterations, such as natural knockouts or deleterious single nucleotide variations, could result in undetectable or low levels of TEX101 protein. We suggested that spermatozoa obtained from such patients could be used as knockout or knockdown models to identify proteins degraded in the absence of TEX101 and discover the functional interactome of TEX101 in humans. Unlike immunoprecipitation approaches to reveal direct and strong physical interactions, profiling of the whole proteome of spermatozoa lacking TEX101 could identify the whole functional interactome including weak and transients interactions. Collectively, such data could support in humans the previously suggested function of TEX101 as a cell-surface chaperone (10).

EXPERIMENTAL PROCEDURES

Study design and statistical rationale

The objectives of this study were to identify potential genomic alterations which could impact levels of TEX101 protein, and verify those levels experimentally in human spermatozoa samples. We suggested that differential proteomic profiling of spermatozoa with such genomic alterations could identify proteins degraded in the absence of TEX101 and thus discover the functional interactome of TEX101 in humans. According to power calculations, differential profiling of spermatozoa of 4 wild-type and 4 rs35033974 homozygous men could identify proteins down-regulated at least 2.4-fold, assuming 80% power, $\alpha = 0.05$, 1.8% coefficient of variation for log₂-transformed LFQ intensity values and a one-tailed t-test. Power calculations were done with G*Power software (version 3.1.7, Heinrich Heine University Dusseldorf). GraphPad Prism (v5.03) was used to generate scatter plots, perform statistical analysis, and calculate the area under the Receiver Operating Characteristic curve (ROC AUC). Non-parametric Mann – Whitney U test was used to compare TEX101 levels in seminal plasma of wild-type and heterozygous men, and *P*-values <0.05 were considered statistically significant.

Study population and sample collection

Semen samples (*N*=386) were collected with informed consent from patients with the approval of the institutional review boards of Mount Sinai Hospital (approval #08-117-E) and University Health Network (#09-0830-AE). Samples were obtained from healthy fertile men before vasectomy and individuals diagnosed with oligospermia or unexplained infertility. Clinical parameters are summarized in **Table 1**. The unexplained infertility group included men who were not able to father a pregnancy after one year of regular unprotected intercourse, with normal sperm concentration of greater than 15 million/mL. After liquefaction, semen samples were centrifuged three times at 13,000 g for 15 min at room temperature. Spermatozoa and SP were separated and stored at -80°C. Samples were analyzed retrospectively. For the differential proteomic analysis, spermatozoa samples were obtained from four men homozygous for *TEX101* c.296G>T variant (rs35033974^{hh}), diagnosed with oligospermia (*N*=2) and unexplained infertility (*N*=2), median age of 29.5 years, sperm concentration 2 – 30 million/mL, and TEX101 concentration in SP 3.5 – 47 ng/mL. Spermatozoa obtained from the age-matched wild-type (WT) fertile men referred for vasectomy (*N*=4, sperm concentration > 15 million/mL and TEX101 concentration in SP of 8,000 – 12,500 ng/mL) were selected as a control group.

Extraction of genomic DNA from spermatozoa and TEX101 genotyping

Genomic DNA was extracted from spermatozoa using QIAamp DNA Mini Kit (Qiagen Inc.). Spermatozoa were washed twice with phosphate-buffered saline (PBS). Cells were lysed in the presence of proteinase K and DNA bound to the membrane was washed and eluted. DNA purity and concentration were measured by spectrophotometer (NanoDrop 8000, Thermo Scientific). Forward (5'-ACAGGACTGAGACAGCCAT-3') and reverse (5'-TCCAGGGTACCTGTGGTCTC-3') primers were designed to amplify a 197 base pair fragment of *TEX101* gene encompassing the rs35033974 polymorphism. PCR was performed with 50 ng of genomic DNA, 1.2 units of Phusion High-Fidelity DNA polymerase (Thermo Scientific) in Phusion HF Buffer, 200 μM deoxynucleoside triphosphates and 0.5 μM primers using mastercycler thermal cycler

(Eppendorf). PCR included an initial denaturation step at 98°C for 1 min, followed by 40 cycles of denaturation at 98°C for 10 s, annealing at 64°C for 30 s and extension at 72°C for 30 s, with a final extension at 72°C for 7 min. PCR products were confirmed with 1.5% agarose gel electrophoresis and purified with QIAquick PCR Purification Kit (Qiagen). Sequencing of PCR products (N=386 patients) was performed by the Centre for Applied Genomic (Hospital for Sick Children, Toronto).

Sample preparation and protein digestion by endopeptidase Glu-C

Spermatozoa pellets were washed twice with PBS, lysed with 0.1% RapiGest SF (Waters, Milford, MA) in 50 mM ammonium bicarbonate and sonicated three times for 30 s. Cell lysates were then centrifuged at 15,000 g for 15 min at 4°C. Total protein in each spermatozoa or SP sample was measured by the bicinchoninic acid assay. Ten µg of total protein per patient sample in 50 mM ammonium bicarbonate were used for protein digestion. RapiGest SF 0.05% with 5 mM dithiothreitol at 65°C for 30 min were used to denature proteins (purified recombinant human rhTEX101 and proteins from spermatozoa and SP) and reduce disulfide bonds. Free thiols were then alkylated with 10 mM iodoacetamide in the dark for 40 min at room temperature. Protein digestion was completed overnight at 37°C in the presence of sequencing grade Glu-C enzyme obtained from Promega (1:20 Glu-C: total protein) and supplemented with 5% acetonitrile to enhance Glu-C activity. Digestion in the presence of ammonium bicarbonate at pH 7.8 ensured specific cleavage after glutamine residues. Trifluoroacetic acid (1%) was then used to inactivate Glu-C and cleave RapiGest SF detergent. Synthetic peptides representing the wild-type (AITIVQHSSPPGLIV*TSYSNYCE) and the G99V variant (AITIVQHSSPPVLIV*TSYSNYCE) forms of TEX101 were labeled with ¹³C5-, ¹⁵N-valine at the residue 102 and were used as internal standards spiked-in after digestion at final concentrations of 100 fmol/µL and 250 fmol/µL, respectively. Digests were desalted, and peptides were extracted by C18 OMIX tips (Varian Inc., Lake Forest, CA). Peptides were eluted into 3 µL of 70% acetonitrile with 0.1% formic acid and analyzed by an EASY-nLC 1000 nanoLC coupled to Q Exactive™ Plus Hybrid Quadrupole-Orbitrap™ Mass Spectrometer (Thermo Fischer Scientific).

Development of Parallel Reaction Monitoring (PRM) assay for WT and G99V variant forms of TEX101 protein

To evaluate Glu-C specificity and efficiency of digestion, rhTEX101 protein was digested and analyzed in the data-dependent discovery mode, which confirmed specific generation of AITIVQHSSPPGLIVTSYSNYCE (m/z=1268.6) and AITIVQHSSPPVLIVTSYSNYCE (m/z=1289.6) peptides. Uniqueness of these peptides in the human proteome was confirmed by Basic Local Alignment Search Tool (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Following that, rhTEX101 and seminal plasma were digested with Glu-C and analyzed in the unscheduled targeted PRM mode. In the final optimized PRM method, heavy isotope-labeled peptide internal standards and an additional endogenous TEX101 peptide TAILATKGCIPE (m/z=637.3) were monitored (**Supplemental Table S1**). A four-step 16-min gradient was used: 20% to 40% of buffer B for 8 min, 40% to 65% for 2 min, 65% to 100% for 2 min, and 100% for 4 min. PRM settings were the following: 3.0 eV in-source CID, 17,500 MS2 resolving power at 200 m/z, 3×10⁶ AGC target, 100 ms injection time, 2.0 m/z isolation window, optimized collision energy at 27 and 100 ms scan times.

Immunocapture-PRM measurements of WT and G99V TEX101

Total TEX101 protein was enriched from SP and spermatozoa using an in-house anti-TEX101 mouse monoclonal antibody 34ED556. Briefly, protein G purified 34ED556 monoclonal antibody was immobilized on NHS-activated Sepharose 4 Fast Flow beads (GE Healthcare). Fifty μL of beads ($\sim 25 \mu\text{g}$ of 34ED556) in 0.1% BSA were incubated overnight at 4°C with seminal plasma or spermatozoa lysate. After binding, beads were washed three times with tris buffer saline (50 mM Tris, 150 mM NaCl, pH 7.5) followed by washing with 50 mM ammonium bicarbonate. Proteins were digested overnight on beads using Glu-C. Supernatants were acidified with 1% TFA. Heavy peptides (200 fmoles of WT and 500 fmoles of G99V) were spiked into each sample after digestion. Digests were desalted, and peptides were measured by PRM assay. Raw files were analyzed with Skyline software (v3.6.0.10493), and the relative abundances of WT or G99V variant TEX101 forms were calculated using the light-to-heavy peptide ratios.

Sample preparation for the differential proteomic analysis

Spermatozoa pellets from eight men (four WT and four rs35033974^{hh}) were lysed with 0.1% RapiGest SF in 50 mM ammonium bicarbonate. Cell lysates were centrifuged at 15,000 g for 15 min at 4°C to remove debris, and total protein concentration was measured using BCA assay. Proteins (225 μg per sample) were denatured, reduced with 5 mM dithiothreitol, alkylated with 10 mM iodoacetamide and digested overnight with trypsin (Sigma-Aldrich) at 37°C .

Strong-cation exchange chromatography fractionation

Off-line SCX fractionation was used to facilitate deep proteome analysis. Tryptic peptides were diluted with mobile phase A (0.26 M formic acid in 10% CAN at pH 2-3) and were loaded onto PolySULFOETHYL A™ column (2.1 mm ID $\times$ 200 mm, 5 μm , 200 Å, The Nest Group Inc., MA). Peptides were separated with a 60-min three step HPLC gradient (Agilent 1100) and eluted at 200 $\mu\text{L}/\text{min}$ with 1 M ammonium formate (0-15% for 5-25 min, 25% at 35 min and 100% at 50 min). Twenty-seven 400 μL fractions were initially collected, but then pooled into 13 fractions based on absorbance profiles.

Protein identification by liquid chromatography - tandem mass spectrometry (LC-MS/MS)

Peptides of each SCX fraction were concentrated with C18 OMIX tips and analyzed by an EASY-nLC 1000 system coupled to a Q Exactive™ Plus mass spectrometer in technical duplicates for each fraction (14, 15). Peptides were separated with a 15 cm C18 analytical column using a 90-min LC gradient at 300 nL/min flow rate. Full MS1 scans (400 to 1500 m/z) were acquired with the Orbitrap analyzer at 70,000 FWHM resolution in the data-dependent mode, followed by 12 data-dependent MS2 scans at 17,500 FWHM. Only +2 and +3 charge states were subjected to MS2 fragmentation.

Data analysis and label-free quantification

XCalibur software (v. 2.0.6; Thermo Fisher Scientific) was utilized to generate raw files. For protein identification and label-free quantification, raw files were analyzed with MaxQuant software (version 1.5.2.8). MaxQuant searches were performed against the non-redundant Human UniprotKB/Swiss-Prot database (HUMAN5640_sProt-072016) at the false discovery rate (FDR) of 1.0%. Search parameters included: trypsin enzyme specificity, two missed cleavages,

minimum peptide length of 7 amino acids, minimum identification of one razor peptide, fixed modification of cysteines by carbamidomethylation and variable modification of methionine oxidation and N-terminal protein acetylation. The mass tolerance was set to 20 ppm for precursor ions and 0.5 Da for fragment ions with top 12 MS/MS peaks per 100 Da. MaxLFQ algorithm facilitated label-free relative quantification of proteins (16). 'ProteinGroups.txt' file was uploaded to Perseus software (version 1.5.5.3) to facilitate statistical analysis (17). Proteins classified as 'Only identified by site', 'Reverse' and 'Contaminants' were filtered out, and LFQ intensities were log2-transformed. Missing LFQ values were imputed with the down shift of 1.8 and distribution width of 0.45 to ensure normal distribution, and average LFQ intensities for two technical replicates were calculated. A two-sample *t*-test with Benjamini-Hochberg false discovery rate-adjusted *p*-values was applied, and 5.0% FDR with calculated constant for variance correction $s_0=0.4$ were used to select proteins differentially expressed in rs35033974^{hh} men. Data was visualized with the volcano plot. Significant up- or down-regulated proteins were filtered for the cell-surface and secreted proteins with the testicular tissue-elevated (tissue-enriched, group enriched and tissue-enhanced) expression according to the Human Protein Atlas, version 13 (1).

Experimental design and rationale for development of Selected Reaction Monitoring (SRM) assays and quantification of candidate proteins

To quantify candidate proteins, we developed and applied Tier 2 SRM assays, as previously described (18-24). Briefly, LC-MS/MS peptide identification data was used to select proteotypic tryptic peptides and develop SRM assays. Choice of peptides was confirmed with the SRM Atlas (www.srmatlas.org). For each protein, peptides with 7-20 amino acids and without missed cleavages were chosen, and heavy isotope-labeled peptide internal standards were synthesized. Several unscheduled 30-min SRM methods were prepared and run with a pool of spermatozoa digest with TSQ Quantiva™ triple quadrupole mass spectrometer (Thermo Scientific). The three most intense transitions were selected for each heavy or light peptide. Finally, 20 heavy and light peptides were scheduled within 2 min intervals during a 30-min gradient in a single multiplex SRM assay (**Supplemental Table S2**). The parameters for SRM assay included: positive polarity, 150 V declustering and 10 V entrance potentials, 300°C ion transfer tube temperature, optimized collision energy values, 20 ms scan time, 0.4 Q1 and 0.7 Q3 FWHM resolutions and 1.5 mTorr Q2 argon pressure. Since one rs35033974 homozygote spermatozoa sample was fully consumed in the discovery experiment, candidate proteins were quantified in three rs35033974 homozygote and four WT spermatozoa samples. Spermatozoa lysates (10 µg protein) were digested by trypsin. TEX101 and DPEP3 internal standards with trypsin-cleavable tags (500 fmoles of AGTETAILATK*-JPTtag and SWSEELQGVLR*-JPTtag, respectively) were added before trypsin digestion, while 8 heavy isotope-labeled peptides without JPT tags were spiked after digestion (500 fmoles each). Light and heavy peptides were monitored with a scheduled 30-min multiplex SRM assay. Each spermatozoa sample was analyzed. Light-to-heavy ratio was used to calculate the accurate relative abundance of each candidate protein.

Western blotting

Protein levels of TEX101, LY6K, ADAM29, and DPEP3 were assessed by Western blot analysis. Twenty µg of total protein from one rs35033974 homozygous and one WT spermatozoa lysate

were loaded onto an SDS-PAGE gel (4% - 15%, Bio-Rad), and transferred onto PVDF membranes (Bio-Rad). After blocking, membranes were incubated overnight at 4°C with rabbit polyclonal antibodies against TEX101 and DPEP3 (HPA041915 and HPA058607, Atlas Antibodies), sheep polyclonal antibody against LY6K (AF6648, R&D Systems), mouse monoclonal antibody against ADAM29 (H00011086-M09, Abnova Corporation) and GAPDH antibody (AM4300, Thermo Fisher Scientific). The membranes were washed then incubated with goat anti-rabbit, donkey anti-sheep, and goat anti-mouse secondary antibody conjugated to horseradish peroxidase (Jackson ImmunoResearch). Proteins were detected with chemiluminescence substrate (GE Healthcare Life Sciences).

Data availability

Raw mass spectrometry shotgun data and MaxQuant output files were deposited to the ProteomeXchange Consortium via PRIDE (www.ebi.ac.uk/pride/archive/login) with the dataset identifier PXD008333 and the following username: reviewer77608@ebi.ac.uk and password: lldaf83v. PRM and SRM raw data were deposited to the Peptide Atlas with the dataset identifier PASS01112 (www.peptideatlas.org/PASS/PASS01112) and the following username: PASS01112 and password: NI5437g. Alternative link is <ftp://PASS01112:NI5437g@ftp.peptideatlas.org>. Processed Skyline files can be downloaded at Panorama Public (<https://panoramaweb.org/3jbthK.url>; email: panorama+diamandis2@proteinms.net, password: jv7Z4#GC).

RESULTS

Database mining for loss-of-function variants of *TEX101* gene

The Exome Aggregation Consortium (<http://exac.broadinstitute.org>), Genome Aggregation (<http://gnomad.broadinstitute.org>) and 1000 Genomes Project (www.internationalgenome.org) databases (25, 26) were examined for the presence of potential loss-of-function variants of the human *TEX101* gene. GnomAD database included genomic variants identified in 138,632 individuals of diverse ethnic background and revealed 166 potential loss-of-function variants of *TEX101* (**Supplemental Table S3**). Protein knockout or truncating variants, such as start loss, stop-gain and frameshift variants, were very rare (minor allele frequencies <0.003%), while some missense variants leading to single amino acid substitutions were much more frequent. One of such missense variants was rs35033974 (allele frequency 8.4%). Interestingly, rs35033974 was predicted as 'deleterious' by Polyphen (27), SIFT (28) and CADD (29) algorithms. The allele frequency of rs35033974 varied among different populations. It was more common in European non-Finnish population (12.4%), less common in Latino (5.2%), Ashkenazi Jewish (5.8%), South Asian (3.3%) and African (2%), and very rare in East Asian population (<0.00001%). Rs35033974 (c. 296 G>T) was localized within exon 4 and resulted in substitution of glycine to valine at position 99 (**Figure 1A**). Alignment of *TEX101* protein sequences suggested that glycine-99 was a conserved residue in 17 of 19 mammals and thus could be intolerant to substitutions (**Supplemental Figure S1**). The high frequency of rs35033974 warranted its identification in our male infertility biobank. Spermatozoa samples obtained from heterozygous (22% genotype frequency in European population) and homozygous (1.6% genotype frequency) men were investigated.

Identification of men heterozygous and homozygous for the rs35033974 variant

Genotypes for rs35033974 variant (c. 296 G>T) were determined in 386 men by amplification of spermatozoa DNA and sequencing analysis (**Table 1** and **Figure 1B**). Four heterozygous individuals (GT, genotype frequency 11%) were identified in the group of pre-vasectomy fertile men ($N=37$). In the group of patients with unexplained male infertility ($N=175$), 23 heterozygous (13%) and 2 (1.1%) homozygous patients (TT) were found. In the group of patients diagnosed with oligospermia ($N=174$), we identified 25 heterozygous (14%) and 2 homozygous (1.1%) patients. Investigation of patients with European ancestry (178 WT, 44 heterozygous and 4 homozygous men) revealed minor allele frequency of 11.5% and was similar to gnomAD frequency of 12.4%. In our cohort of European men, the minor allele frequency was higher for patients with unexplained infertility and oligospermia (12.7%) than infertile men pre-vasectomy (5.4%), but was not significant (Fisher's exact test $P=0.15$).

Impact of rs35033974 on TEX101 protein concentration in seminal plasma

We previously measured by ELISA concentration of total TEX101 protein in seminal plasma of 805 men (8). Cross-checking revealed that concentration of total TEX101 in seminal plasma of four rs35033974^{hh} homozygous patients was extremely low (3.5 to 47 ng/mL), despite of their medium-to-high sperm concentration (2 to 30 mln/mL; **Figure 1C**). Here, we also examined possible associations between rs35033974 heterozygous status and TEX101 levels in seminal plasma or sperm concentration in semen. No significant difference for sperm concentration was found for WT versus heterozygous European population (MWU P -value =0.94). However, levels of TEX101 in seminal plasma were significantly lower for rs35033974 heterozygous men (median 390 ng/mL, MWU P =0.0005, N =40), as compared to WT men of European population (median 1,949 ng/mL, N =145; **Figure 1D**). Since TEX101 concentration in seminal plasma may correlate with the number of spermatozoa in semen, we also investigated normalized TEX101 concentration. Thus, seminal plasma levels of TEX101 normalized by sperm concentration were significantly lower (4-fold change, MWU P <0.0001) for heterozygous men (median 40,738 attograms/cell, N =40), as compared to WT men (median 164,085 attograms/cell, N =145).

Rs35033974 results in degradation of G99V TEX101 protein

To monitor the WT and G99V variant forms of TEX101 protein in spermatozoa lysate and SP, we opted to develop a targeted mass spectrometry assay (19, 30, 31). Since TEX101 digestion by trypsin generated a 38-amino acid peptide not suitable for bottom-up proteomic measurements, we explored alternative proteases, such as endopeptidase Glu-C (cleaves peptide sequences after aspartic and glutamic acids) and neutrophil elastase (cleaves after valine and alanine). Endopeptidase Glu-C was found the most suitable enzyme. Following optimization of Glu-C digestion protocol, we developed a PRM assay to monitor WT and G99V TEX101 peptides, as well as an additional endogenous "control" peptide which represented total TEX101. Sensitivity of PRM assay, however, was not sufficient to measure low levels of TEX101 in seminal plasma (<1.5 µg/mL).

To improve assay sensitivity, we developed an immuno-PRM assay based on the immunoenrichment of TEX101 by our in-house anti-TEX101 mouse monoclonal antibody 34ED556 coupled to sepharose beads (**Figure 2A**). Using immuno-PRM assay, we were able to identify barely detectable levels of G99V variant protein in one homozygous patient, while the other two homozygous patients had undetectable levels of G99V protein. We also selected one heterozygous patient with a very high concentration of TEX101 in seminal plasma (16.1 µg/mL), and were able to measure levels of both WT and G99V variant forms in his spermatozoa (**Figure 2B**). Interestingly, the abundance of a G99V variant form was substantially lower, as compared to the WT form. Using heavy-to-light ratios of both peptides, we estimated that the abundance of

the G99V form was ~97% lower than predicted, assuming equal expression of both alleles.

To explain this phenomenon, we suggested that the G99V variant form of TEX101 protein could be misfolded, aggregated and destroyed through proteasome or autophagosome degradation. A large residue of valine at position 99 could introduce substantial steric constraints, eliminate a proline-glycine beta-turn and destabilize beta-sheets in the proximity of G99V (**Supplemental Figure S2**). Interestingly, TANGO algorithm (32) for prediction of cross-beta aggregation in unfolded proteins predicted a significant impact of G99V substitution. Predicted cross-beta aggregation values changed from 3.6 for G99 to 83.8 for G99V (the average values for all residues of the mature WT TEX101 was 3.8).

Global proteomic profiling revealed testis-specific proteins down-regulated in rs35033974^{hh} spermatozoa

Spermatozoa obtained from four rs35033974^{hh} men were considered as a *TEX101* functional knockdown model. Taking into account degradation of ADAM3 proteins in spermatozoa of *Tex101* knockout mice (10), we hypothesized that TEX101-dependent human proteins would be degraded in rs35033974^{hh} spermatozoa and could be identified by differential proteomic profiling. Unlike immunoprecipitation approaches to identify only direct and strong physical interactions, differential profiling of the whole proteomes of WT versus rs35033974^{hh} spermatozoa could identify all interactions which deteriorated in the absence of TEX101 (strong, weak, direct, transient and indirect).

We thus performed a global proteomic analysis of four WT and four rs35033974^{hh} spermatozoa (**Figure 3A**). To achieve deep proteome coverage, peptides were subjected to the off-line fractionation by strong-cation exchange chromatography followed by the online reversed-phase liquid chromatography-mass spectrometry detection. As a result, MaxQuant analysis identified and quantified 83,984 unique peptides and 8,046 protein groups with FDR≤1.0% (**Supplemental Table S4**). Of 189 differentially regulated proteins (FDR≤5.0% and $s_0=0.4$), 96 were down-regulated and 93 were up-regulated (**Figure 3B and Supplemental Table S5**). Filtering of these candidates for testis specificity using Human Protein Atlas data revealed 55 down-regulated but only 4 up-regulated proteins (potential false-positives). Thus, many more testis-specific proteins were affected by TEX101 loss. Additional filtering for cell-surface and secreted proteins using NextProt database revealed 8 down-regulated cell-surface and 9 secreted proteins, but no any up-regulated proteins (**Figure 3B**).

Here, we also hypothesized that differential proteomic analysis of rs35033974^{hh} spermatozoa could reveal functional orthologs of mouse ADAMs 3-6 proteins destroyed in *Tex101* knockout mice (10, 12). In our spermatozoa proteome, we identified all seven human testis-specific ADAM proteins, of which three proteins with the adhesion activity (ADAM18, ADAM29 and ADAM32) could be potential orthologs of mouse ADAM 3-6

proteins (33). It was only ADAM29 protein which levels were lower in rs35033974^{hh} spermatozoa (**Figure 3B**). Even though ADAM29 did not pass our cut-off criteria of the global differential analysis, it was down-regulated 2.5-fold (BH-adjusted t-test $P=0.003$). We thus decided to further evaluate ADAM29 protein by SRM and Western blotting.

In this work, we identified and quantified one of the largest proteomes of human spermatozoa (8,046 protein groups representing 8,473 unique gene names and 8,431 UniProt protein IDs). Of these, 7,156 protein groups and 7,573 unique Uniprot IDs were quantified with 2 or more unique peptides. Interestingly, of 2,186 proteins currently defined as “missing” by NextProt (v2.14.0), we identified 127 proteins with 2 unique peptides (101 with previous evidence only at the transcript level, 21 inferred from homology and 5 predicted). For example, we discovered in our dataset 7 testis-elevated “missing” proteins (ANKRD60, C12orf42, LRRC63, CCDC74B, FAM47C, SPATA31A1 and TTLL8) which were identified with 2 unique peptides (≥ 9 amino acids) and could represent true proteins according to NextProt criteria. To conclude, our spermatozoa proteome could be used as a resource to identify and update currently “missing” testis-specific proteins.

Evaluation of candidates by SRM and western blotting

ADAM29 and seven testis-specific cell-surface and secreted proteins involved in sperm migration (34), zona pellucida binding and penetration (35-37) and sperm-oocyte fusion (38, 39) were selected for SRM analysis. As a result of SRM measurements, 7 of 8 proteins were down-regulated in rs35033974^{hh} spermatozoa, while levels of DPEP3 (a testis-specific TEX101-interacting protein not affected by *Tex101* knockout in mice (11)), were not changed (**Figure 3B and Figure 4A and Table 2**).

Interestingly, the top down-regulated protein in rs35033974^{hh} spermatozoa was LY6K (14-fold change; *t*-test P -value=0.03). LY6K (Lymphocyte antigen 6K) is a testis-specific GPI-anchored protein localized at the cell surface of testicular germ cells, with a similar expression pattern to TEX101 based on immunohistochemistry data available at the Human Protein Atlas. It was previously demonstrated in mice that LY6K physically interacted with TEX101, but disappeared from the cell surface in the absence of TEX101 protein in *Tex101* knockout mice (11, 34). Finally, we evaluated LY6K, ADAM29 and TEX101 proteins by western blotting. Levels of these proteins were undetectable in rs35033974^{hh} spermatozoa, while expression of DPEP3 monomers (75 kDa) and dimers (150 kDa) was found at normal levels (**Figure 4B**).

DISCUSSION

It is estimated that 30% of infertile men are diagnosed with unexplained infertility, often with normal semen parameters of sperm count and motility (40). The routine testing for male infertility is currently limited to analysis of sperm concentration, motility and morphology (41), as well analysis of few genetic abnormalities including Y chromosome deletions, CFTR mutations and sperm DNA fragmentation for specific indications (42). Many more genetic abnormalities associated with male infertility have been identified (43-47), however, improved diagnostics of male infertility remains an unmet need. Elucidation of functional roles of testis-specific proteins may not only advance biology of human reproduction, but also provide novel diagnostic markers, particularly for male infertility subtypes which cannot currently be explained with routine diagnostics.

To identify genes essential for male fertility, approximately 500 knockout mouse models have been investigated to date. Some models resulted in male infertility with severe defects in spermatogenesis, spermiogenesis and sperm maturation, or infertility with no apparent defects in motility, morphology and sperm count (*Ace*, *Adam2* and *Adam3*) (48, 49). Other models revealed sub-fertile phenotypes with reduced acrosome reaction and capacitation (*Acr*, *Pcsk4*), decreased motility (*Smcp*), delayed cumulus-oocyte complex dispersion (*Spam1*), weak zona pellucida binding (*Press21*) and reduced sperm-oocyte fusion (*Crisp1*) (50). Surprisingly, knockout studies in mice also revealed numerous sperm cell surface proteins not essential for fertilization *in vivo*, thus suggesting compensatory mechanisms and multi-factorial nature of infertility (50).

Recently, the Human Protein Atlas profiled testis-specific genes and proteins, and identified 1,079 genes with more than five-fold higher mRNA levels in testis as compared to all other human tissues (1). Due to their exclusive expression in testis, these proteins may be essential for spermatogenesis, remodeling of sperm surface proteome, sperm transit and sperm-oocyte fusion (51). However, lack of cell lines expressing testis-specific proteins was the major bottleneck to study the molecular function of human testis-specific proteins *in vitro*. Likewise, primary human germ cells isolated from orchiectomy samples could not be maintained in long-term cultures and studied with gene knockout or knockdown approaches.

Identification of natural 'human knockouts' with homozygous loss-of-function mutations provided an alternative approach to study functional and pathological roles of human proteins (52). The most valuable mutations for such studies included protein truncating variants due to stop gain or frameshift mutations, or single amino acid variants leading to the loss of activity or to protein misfolding followed by proteasomal degradation. In this work, we hypothesized that spermatozoa of men with natural knockouts or functional knockdowns of testis-specific genes could emerge as valuable models to study functional and pathological roles of human testis-specific proteins.

Recent population-scale studies of genetic variation (25, 53, 54) discovered numerous protein truncating variants or single amino acid variants in humans, and provided their accurate frequencies in different ethnic groups (55). For example, the Genome Aggregation Database (www.gnomad.broadinstitute.org) contained the whole-exome and whole-genome sequencing data for 138,632 individuals with loss-of-function mutations and predicted deleterious impact of missense variants with Polyphen and SIFT algorithms (25). Variants present with the minor allele frequencies $\geq 0.3\%$ and 5.5% could be of particular interest since such frequencies ensure identification of at least 3 heterozygous or homozygous individuals, respectively, per 1,000 individuals in the general population. It is expected that frequencies of pathogenic variants will be further enriched in patient cohorts, such as men referred to fertility clinics. Thus, the impact of truncating or single amino acid variants on the expression or activity of corresponding testis-specific proteins could be investigated experimentally in clinical samples, such as spermatozoa or testicular tissues. Since immunoassays are not available for the majority of testis-specific proteins and also typically cannot detect single amino acid substitutions, mass spectrometry-based approaches remain the only tools to measure the expression of variant proteins (56, 57).

In this study, we focused on a testis-specific protein TEX101 which we previously identified and validated as a seminal plasma biomarker for the differential diagnosis of azoospermia and male infertility (58-60). TEX101 was previously shown to be essential for the production of fertilization-competent spermatozoa in mice (10, 12, 61-63). Targeted knockouts of *Tex101* gene resulted in absolute male sterility, with no other phenotypic abnormalities and normal sperm concentration and morphology (10, 11). The presence of TEX101 protein was shown to be crucial for the trafficking, maturation or localization of four proteins of the ADAM family (ADAM3, ADAM4, ADAM5 and ADAM6) essential for fertilization in mice (10-12). However, mouse data could not be translated into human studies since human ADAM3, ADAM5 and ADAM6 were non-coding pseudogenes, while ADAM4 was not present in the human genome (13).

Here, we searched population-scale genomic databases for the loss-of-function protein truncating and single amino acid variants of human *TEX101* gene, identified rs35033974 variant and discovered substantially lower levels of the G99V variant form of TEX101 protein in spermatozoa of heterozygous and homozygous men. We then hypothesized that rs35033974^{hh} spermatozoa could be used as a knockdown model to study the molecular function of human TEX101, discover its functional interactome and identify orthologs of mouse ADAMs 3-6. Our proteomic approach was designed to identify the TEX101 functional interactome, such as proteins co-degraded together with G99V TEX101 in rs35033974^{hh} spermatozoa. We hypothesized that some of these interactions could be weak or transient and thus missed by typical co-immunoprecipitation approaches.

As a result, we identified and verified 8 cell-surface and 9 secreted testis-specific proteins which were significantly down-regulated in rs35033974^{hh} spermatozoa. It should be emphasized that identification of LY6K protein as a top candidate suggested the robustness of our experimental protocol. Indeed, previous studies in mice revealed that LY6K physically interacted with TEX101 and disappeared from the spermatozoa surface in *Tex101* knockout mice (62).

Similar to TEX101, LY6K is GPI-anchored cell-surface protein expressed by testicular germ cells and is partially shed into seminal plasma during sperm maturation (34). Interaction of TEX101 with LY6K in mice was shown to be crucial for proper trafficking and post-translational processing of LY6K (11). *Tex101*^{-/-} or *Ly6k*^{-/-} mice were infertile due to compromised migration of sperm in the oviduct (10, 11, 34). TEX101-LY6K complex facilitated proper processing of ADAM3 protein. Interestingly, levels of TEX101 and LY6K proteins on the surface of spermatozoa, but not levels of intracellular mRNA transcripts, were mutually dependent. Thus, LY6K protein quickly degraded in *Tex101*^{-/-} mice, and vice versa (11). Interestingly, another cell-surface GPI-anchored protein, a testis-specific dipeptidase DPEP3, also formed a physical complex with TEX101 (63), but DPEP3 levels were not affected in *Tex101*^{-/-} mice (11). Thus, our data on human TEX101, LY6K and DPEP3 in WT and rs35033974^{hh} spermatozoa (**Figure 4**) confirmed previous observations in mice.

Global proteomic profiling of spermatozoa from four WT and four rs35033974^{hh} men identified eight testis-specific ADAM proteins with adhesion (ADAM2, ADAM18, ADAM29, ADAM32) and metalloprotease (ADAM20, ADAM21, ADAM28, ADAM30) activities (33). Interestingly, it was only ADAM29 which levels decreased in rs35033974^{hh} men, as discovered by global proteomic profiling (2.5-fold, *P*=0.003) and verified by SRM and western blotting. Since molecular function of human ADAM29 protein has never been previously reported, we suggest that ADAM29 protein should be further investigated as one of the potential functional orthologs of mouse ADAM 3-6 proteins.

We also suggested here that very low levels of G99V variant form in heterozygous and homozygous spermatozoa (~3% of predicted) could be the result of misfolding and subsequent proteasomal degradation (64-66). Similar impact was previously observed for misfolded CFTR (67) and some GPI-anchored proteins (68). Introduction of a large hydrophobic valine residue at position 99 within the PPGL beta-turn could result in substantial steric constraints, destabilization of proximal beta-sheets and subsequent protein aggregation (69, 70). Significant impact of G99V substitution on beta-aggregation was in agreement with TANGO calculations (32). The predicted TANGO impact was the most deleterious for substitutions of G99 with hydrophobic residues such as valine, isoleucine and phenylalanine. Glycine at position 99 was also found highly conserved in mammals (**Supplemental Figure S1**).

In this work, we identified and quantified one of the largest proteomes of human spermatozoa (8,046 protein groups representing 8,473 unique gene names and 8,431

UniProt protein IDs). At least 7 testis-elevated proteins (ANKRD60, C12orf42, LRRC63, CCDC74B, FAM47C, SPATA31A1 and TTLL8) identified in our work were previously considered “missing proteins” in NextProt, UniProt and Human Protein Atlas databases. Our deep proteome of spermatozoa could be used as a resource to update currently “missing” testis-expressed proteins.

Finally, we speculated on possible deleterious impact of rs35033974 allele on male fertility. Even though TEX101 protein levels were found significantly lower in rs35033974 heterozygous men, 70% of heterozygous men still had TEX101 levels in seminal plasma above 65 ng/mL, a cut-off level of male infertility which we previously established (8). Due to its relatively high allele frequency in the general population (12.4% in Europeans), rs35033974 is unlikely a severe pathogenic factor leading to male sterility. In addition, if rs35033974 would be a recessive pathogenic variant with a high penetrance, the number of homozygous men in population would be substantially lower than that calculated using the Hardy-Weinberg equilibrium. We found no substantial deviation from the equilibrium for the European population in gnomAD ($N=63,332$; 1,011 homozygotes identified versus 982 calculated). According to the 1000 Genomes Project, the rs35033974 variant was present with the similar allele frequencies in males and females. Interestingly, 1000 Genomes Project data (26) included 9 homozygous men, of which 5 had biological children (see examples in **Supplemental Figure S3**). Since TEX101 levels were found significantly lower in heterozygous and homozygous men, rs35033974 status may be considered in the use of TEX101 protein as a male infertility biomarker. In addition, patients with different ethnic backgrounds may have different cut-off values for TEX101, with lower cut-offs for European men, of which 21.8% are heterozygous and 1.6% are homozygous.

There may be several possibilities to explain the discrepancy between molecular and clinical data on TEX101 protein: (i) unlike mouse TEX101, human TEX101 protein and TEX101-LY6K interaction may not be essential for sperm maturation, ADAM processing and fertilization; (ii) low levels of TEX101 protein in rs35033974^{hh} patients may be compensated by alternative cell-surface chaperons; (iii) rs35033974 is deleterious for protein structure, however, unlike mouse *Tex101*, human *TEX101* could be another non-essential gene (71, 72). Future studies should investigate if this highly-frequent variant (1.6% homozygous genotype frequency in European population) predisposes males to infertility and becomes pathogenic in combination with other factors, for example, lowered sperm concentration in semen.

It should be noted that our study had the following limitations: (i) even though label-free quantification using MaxQuant algorithm is recognized as an accurate proteome-wide quantification approach (16), its variability may still be relatively high, so all candidates should be verified by orthogonal assays, such as SRM or western blotting; (ii) global proteomic quantification of a very large number of proteins (8,046) and FDR-based cut-offs could result in many false-positive (for example, intracellular non-testis-

specific proteins) and false-negative candidates (ADAM29 could be such a false-negative candidate); (iii) while the results of our global proteomic quantification were confirmed by SRM assays in three rs35033974^{hh} and four WT patients, and by western blotting in single rs35033974^{hh} and WT patients, we could not recruit additional rs35033974^{hh} patients for further validation of our data; (iv) fertility status of our four rs35033974^{hh} patients is not known. One rs35033974^{hh} patient proceeded to infertility treatment by intrauterine insemination, but the treatment of that couple was not successful (female factor, however, was not excluded).

To conclude, we presented the first human study to investigate the possible functional role of TEX101 protein as a cell-surface chaperone, and identified degradation of LY6K, as well as additional 7 cell-surface testis-specific proteins, in the absence of TEX101. Spermatozoa of rs35033974^{hh} men may be used as a unique model to elucidate further details on the role of human TEX101 which will advance male reproduction biology. Since TEX101 seminal plasma levels were found significantly lower in heterozygous than in wild-type men, rs35033974 status could be considered in TEX101 diagnostics. Further studies on rs35033974 and its association with male infertility diagnosis or prognosis will be required. Our work may serve as a concept for future studies on functional effects of natural knockouts or knockdowns in humans and could lead to development of novel male infertility diagnostics (73, 74) and discovery of targets for male contraceptives (75). Presented approach may also facilitate verification of the essential and non-essential testis-specific genes and proteins, which will advance biology of human reproduction.

Acknowledgements: We thank Ihor Batruch for assistance with mass spectrometry, and Susan Lau for coordinating collection and storage of clinical samples.

Funding: This work was supported by the Canadian Institute of Health Research Proof of Principle Program - Phase I grants (#303100 and 355146) to K.J., A.P.D. and E.P.D., and Physicians Services Incorporated Foundation Health Research Grant to K.J., A.P.D. and E.P.D.

Author contributions: A.P.D., E.P.D. and K.J. designed the research project; C.S. performed the experiments; C.S. and A.P.D. analyzed the data. D.K. produced anti-TEX101 mouse monoclonal antibodies, and K.J. provided clinical samples and clinical expertise. C.S. and A.P.D. wrote the manuscript, and all authors contributed to the revision of the manuscript.

Competing interests: K.J., E.P.D. and A.P.D. were granted the United States Patent 9040464 "Markers of the male urogenital tract." Other authors declare that they have no competing interests.

Supplemental materials: This article contains supplemental materials.

REFERENCES

1. Uhlen, M., Fagerberg, L., Hallstrom, B. M., et al. (2015) Proteomics. Tissue-based map of the human proteome. *Science* 347, 1260419
2. Nuti, F., and Krausz, C. (2008) Gene polymorphisms/mutations relevant to abnormal spermatogenesis. *Reprod Biomed Online* 16, 504-513
3. Eberhard, J., Stahl, O., Giwercman, Y., et al. (2004) Impact of therapy and androgen receptor polymorphism on sperm concentration in men treated for testicular germ cell cancer: a longitudinal study. *Hum Reprod* 19, 1418-1425
4. Tuttelmann, F., Rajpert-De Meyts, E., Nieschlag, E., et al. (2007) Gene polymorphisms and male infertility--a meta-analysis and literature review. *Reprod Biomed Online* 15, 643-658
5. Drabovich, A. P., Dimitromanolakis, A., Saraon, P., et al. (2013) Differential diagnosis of azoospermia with proteomic biomarkers ECM1 and TEX101 quantified in seminal plasma. *Sci Transl Med* 5, 212ra160
6. Drabovich, A. P., Jarvi, K., and Diamandis, E. P. (2011) Verification of male infertility biomarkers in seminal plasma by multiplex selected reaction monitoring assay. *Mol Cell Proteomics* 10, M110.004127
7. Korbakis, D., Brinc, D., Schiza, C., et al. (2015) Immunocapture-Selected Reaction Monitoring Screening Facilitates the Development of ELISA for the Measurement of Native TEX101 in Biological Fluids. *Mol Cell Proteomics* 14, 1517-1526
8. Korbakis, D., Schiza, C., Brinc, D., et al. (2017) Preclinical evaluation of a TEX101 protein ELISA test for the differential diagnosis of male infertility. *BMC Med* 15, 60
9. Schiza, C. (2017) Preclinical Evaluation of TEX101 Protein as a Male Infertility Biomarker and Identification of its Functional Interactome. pp. 1-196, University of Toronto, Toronto
10. Fujihara, Y., Tokuhira, K., Muro, Y., et al. (2013) Expression of TEX101, regulated by ACE, is essential for the production of fertile mouse spermatozoa. *Proc Natl Acad Sci USA* 110, 8111-8116
11. Endo, S., Yoshitake, H., Tsukamoto, H., et al. (2016) TEX101, a glycoprotein essential for sperm fertility, is required for stable expression of Ly6k on testicular germ cells. *Sci Rep* 6, 23616
12. Li, W., Guo, X. J., Teng, F., et al. (2013) Tex101 is essential for male fertility by affecting sperm migration into the oviduct in mice. *J Mol Cell Biol* 5, 345-347
13. Cho, C. (2012) Testicular and epididymal ADAMs: expression and function during fertilization. *Nat Rev Urol* 9, 550-560
14. Begcevic, I., Brinc, D., Drabovich, A. P., et al. (2016) Identification of brain-enriched proteins in the cerebrospinal fluid proteome by LC-MS/MS profiling and mining of the Human Protein Atlas. *Clin Proteomics* 13, 11
15. Cho, C. K., Drabovich, A. P., Karagiannis, G. S., et al. (2013) Quantitative proteomic analysis of amniocytes reveals potentially dysregulated molecular networks in Down syndrome. *Clin Proteomics* 10, 2
16. Cox, J., Hein, M. Y., Lubner, C. A., et al. (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* 13, 2513-2526
17. Cox, J., and Mann, M. (2012) 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. *BMC Bioinformatics* 13 Suppl 16, S12
18. Drabovich, A. P., Pavlou, M. P., Schiza, C., et al. (2016) Dynamics of Protein Expression Reveals Primary Targets and Secondary Messengers of Estrogen Receptor Alpha Signaling in MCF-7 Breast Cancer Cells. *Mol Cell Proteomics* 15, 2093-2107

19. Karakosta, T. D., Soosaipillai, A., Diamandis, E. P., et al. (2016) Quantification of Human Kallikrein-Related Peptidases in Biological Fluids by Multiplatform Targeted Mass Spectrometry Assays. *Mol Cell Proteomics* 15, 2863-2876
20. Drabovich, A. P., Pavlou, M. P., Dimitromanolakis, A., et al. (2012) Quantitative analysis of energy metabolic pathways in MCF-7 breast cancer cells by selected reaction monitoring assay. *Mol Cell Proteomics* 11, 422-434
21. Martinez-Morillo, E., Nielsen, H. M., Batruch, I., et al. (2014) Assessment of peptide chemical modifications on the development of an accurate and precise multiplex selected reaction monitoring assay for apolipoprotein e isoforms. *J Proteome Res* 13, 1077-1087
22. Drabovich, A. P., and Diamandis, E. P. (2010) Combinatorial peptide libraries facilitate development of multiple reaction monitoring assays for low-abundance proteins. *J Proteome Res* 9, 1236-1245
23. Prakash, A., Rezai, T., Krastins, B., et al. (2010) Platform for establishing interlaboratory reproducibility of selected reaction monitoring-based mass spectrometry peptide assays. *J Proteome Res* 9, 6678-6688
24. Prakash, A., Rezai, T., Krastins, B., et al. (2012) Interlaboratory reproducibility of selective reaction monitoring assays using multiple upfront analyte enrichment strategies. *J Proteome Res* 11, 3986-3995
25. Lek, M., Karczewski, K. J., Minikel, E. V., et al. (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285-291
26. Genomes Project, C., Auton, A., Brooks, L. D., et al. (2015) A global reference for human genetic variation. *Nature* 526, 68-74
27. Adzhubei, I. A., Schmidt, S., Peshkin, L., et al. (2010) A method and server for predicting damaging missense mutations. *Nat Methods* 7, 248-249
28. Sim, N. L., Kumar, P., Hu, J., et al. (2012) SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res* 40, W452-457
29. Itan, Y., Shang, L., Boisson, B., et al. (2016) The mutation significance cutoff: gene-level thresholds for variant predictions. *Nat Methods* 13, 109-110
30. Martinez-Morillo, E., Cho, C. K., Drabovich, A. P., et al. (2012) Development of a multiplex selected reaction monitoring assay for quantification of biochemical markers of down syndrome in amniotic fluid samples. *J Proteome Res* 11, 3880-3887
31. Begcevic, I., Brinc, D., Dukic, L., et al. (2018) Targeted mass spectrometry-based assays for relative quantification of thirty brain-related proteins and their clinical applications. *J Proteome Res*
32. Fernandez-Escamilla, A. M., Rousseau, F., Schymkowitz, J., et al. (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol* 22, 1302-1306
33. Primakoff, P., and Myles, D. G. (2000) The ADAM gene family: surface proteins with adhesion and protease activity. *Trends Genet* 16, 83-87
34. Fujihara, Y., Okabe, M., and Ikawa, M. (2014) GPI-anchored protein complex, LY6K/TEX101, is required for sperm migration into the oviduct and male fertility in mice. *Biol Reprod* 90, 60
35. Baba, D., Kashiwabara, S., Honda, A., et al. (2002) Mouse sperm lacking cell surface hyaluronidase PH-20 can pass through the layer of cumulus cells and fertilize the egg. *J Biol Chem* 277, 30310-30314
36. Lin, Y. N., Roy, A., Yan, W., et al. (2007) Loss of zona pellucida binding proteins in the acrosomal matrix disrupts acrosome biogenesis and sperm morphogenesis. *Mol Cell Biol* 27, 6794-6805
37. Mandal, A., Klotz, K. L., Shetty, J., et al. (2003) SLLP1, a unique, intra-acrosomal, non-bacteriolytic, c lysozyme-like protein of human spermatozoa. *Biol Reprod* 68, 1525-1537
38. Adham, I. M., Nayernia, K., and Engel, W. (1997) Spermatozoa lacking acrosin protein show delayed fertilization. *Mol Reprod Dev* 46, 370-376

39. Inoue, N., Ikawa, M., Isotani, A., et al. (2005) The immunoglobulin superfamily protein Izumo is required for sperm to fuse with eggs. *Nature* 434, 234-238
40. Pierik, F. H., Van Ginneken, A. M., Dohle, G. R., et al. (2000) The advantages of standardized evaluation of male infertility. *Int J Androl* 23, 340-346
41. World Health Organization, D. o. R. H. a. R. (2010) WHO laboratory manual for the examination and processing of human semen. *ISBN 978 92 4 154778 9*
42. Foresta, C., Ferlin, A., Gianaroli, L., et al. (2002) Guidelines for the appropriate use of genetic tests in infertile couples. *Eur J Hum Genet* 10, 303-312
43. Yatsenko, A. N., O'Neil, D. S., Roy, A., et al. (2012) Association of mutations in the zona pellucida binding protein 1 (ZPBP1) gene with abnormal sperm head morphology in infertile men. *Mol Hum Reprod* 18, 14-21
44. Ramasamy, R., Bakircioglu, M. E., Cengiz, C., et al. (2015) Whole-exome sequencing identifies novel homozygous mutation in NPAS2 in family with nonobstructive azoospermia. *Fertil Steril* 104, 286-291
45. Choi, Y., Jeon, S., Choi, M., et al. (2010) Mutations in SOHLH1 gene associate with nonobstructive azoospermia. *Hum Mutat* 31, 788-793
46. Aston, K. I., and Carrell, D. T. (2009) Genome-wide study of single-nucleotide polymorphisms associated with azoospermia and severe oligozoospermia. *J Androl* 30, 711-725
47. Aston, K. I., Krausz, C., Laface, I., et al. (2010) Evaluation of 172 candidate polymorphisms for association with oligozoospermia or azoospermia in a large cohort of men of European descent. *Hum Reprod* 25, 1383-1397
48. Matzuk, M. M., and Lamb, D. J. (2002) Genetic dissection of mammalian fertility pathways. *Nat Cell Biol* 4 Suppl, s41-49
49. Matzuk, M. M., and Lamb, D. J. (2008) The biology of infertility: research advances and clinical challenges. *Nat Med* 14, 1197-1213
50. Ikawa, M., Inoue, N., Benham, A. M., et al. (2010) Fertilization: a sperm's journey to and interaction with the oocyte. *J Clin Invest* 120, 984-994
51. Djureinovic, D., Fagerberg, L., Hallstrom, B., et al. (2014) The human testis-specific proteome defined by transcriptomics and antibody-based profiling. *Mol Hum Reprod* 20, 476-488
52. Saleheen, D., Natarajan, P., Armean, I. M., et al. (2017) Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. *Nature* 544, 235-239
53. Fu, W., O'Connor, T. D., Jun, G., et al. (2013) Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* 493, 216-220
54. Consortium, T. G. P. (2015) A global reference for human genetic variation. *Nature*, 68-74
55. Alkuraya, F. S. (2015) Natural human knockouts and the era of genotype to phenotype. *Genome Med* 7, 48
56. Li, J., Su, Z., Ma, Z. Q., et al. (2011) A bioinformatics workflow for variant peptide detection in shotgun proteomics. *Mol Cell Proteomics* 10, M110 006536
57. Dimitrakopoulos, L., Prassas, I., Diamandis, E. P., et al. (2016) Proteogenomics: Opportunities and Caveats. *Clin Chem* 62, 551-557
58. Schiza, C. G., Jarv, K., Diamandis, E. P., et al. (2014) An Emerging Role of TEX101 Protein as a Male Infertility Biomarker. *EJIFCC* 25, 9-26
59. Drabovich, A. P., Saraon, P., Jarvi, K., et al. (2014) Seminal plasma as a diagnostic fluid for male reproductive system disorders. *Nat Rev Urol* 11, 278-288
60. Bieniek, J. M., Drabovich, A. P., and Lo, K. C. (2016) Seminal biomarkers for the evaluation of male infertility. *Asian J Androl* 18, 426-433
61. Tsukamoto, H., Yoshitake, H., Mori, M., et al. (2006) Testicular proteins associated with the germ cell-marker, TEX101: involvement of cellubrevin in TEX101-trafficking to the cell surface during spermatogenesis. *Biochem Biophys Res Commun* 345, 229-238

62. Yoshitake, H., Tsukamoto, H., Maruyama-Fukushima, M., et al. (2008) TEX101, a germ cell-marker glycoprotein, is associated with lymphocyte antigen 6 complex locus k within the mouse testis. *Biochem Biophys Res Commun* 372, 277-282
63. Yoshitake, H., Yanagida, M., Maruyama, M., et al. (2011) Molecular characterization and expression of dipeptidase 3, a testis-specific membrane-bound dipeptidase: complex formation with TEX101, a germ-cell-specific antigen in the mouse testis. *J Reprod Immunol* 90, 202-213
64. Loo, T. W., and Clarke, D. M. (1994) Functional consequences of glycine mutations in the predicted cytoplasmic loops of P-glycoprotein. *J Biol Chem* 269, 7243-7248
65. Wang, Y. J., Tayo, B. O., Bandyopadhyay, A., et al. (2014) The association of the vanin-1 N131S variant with blood pressure is mediated by endoplasmic reticulum-associated degradation and loss of function. *PLoS Genet* 10, e1004641
66. Vembar, S. S., and Brodsky, J. L. (2008) One step at a time: endoplasmic reticulum-associated degradation. *Nat Rev Mol Cell Biol* 9, 944-957
67. Ward, C. L., Omura, S., and Kopito, R. R. (1995) Degradation of CFTR by the ubiquitin-proteasome pathway. *Cell* 83, 121-127
68. Mayor, S., and Riezman, H. (2004) Sorting GPI-anchored proteins. *Nat Rev Mol Cell Biol* 5, 110-120
69. Betts M. J., R. R. B. (2003) Amino Acid Properties and Consequences of Substitutions. *Bioinformatics for Geneticists*.
70. Sunyaev, S., Lathe, W., 3rd, and Bork, P. (2001) Integration of genome data and protein structures: prediction of protein folds, protein interactions and "molecular phenotypes" of single nucleotide polymorphisms. *Curr Opin Struct Biol* 11, 125-130
71. Schumacher, J., Zischler, H., and Herlyn, H. (2017) Effects of different kinds of essentiality on sequence evolution of human testis proteins. *Sci Rep* 7, 43534
72. Wang, T., Birsoy, K., Hughes, N. W., et al. (2015) Identification and characterization of essential genes in the human genome. *Science* 350, 1096-1101
73. Drabovich, A. P., Martinez-Morillo, E., and Diamandis, E. P. (2015) Toward an integrated pipeline for protein biomarker development. *Biochim Biophys Acta* 1854, 677-686
74. Drabovich, A. P., Pavlou, M. P., Batruch, I., et al. (2013) Proteomic and Mass Spectrometry Technologies for Biomarker Discovery. In: Issaq, H. J., and Veenstra, T. D., eds. *Proteomic and Metabolomic Approaches to Biomarker Discovery*, pp. 17-37, Academic Press (Elsevier), Waltham, MA
75. Amory, J. K. (2016) Male contraception. *Fertil Steril* 106, 1303-1309

TABLES

Table 1. Clinicopathological variables of 386 patients.

Clinical parameters	N	%
Number of patients	386	100
Median age [range]	41 [19 – 63]	
Ethnic background of patients		
African – Canadian	15	3.9
Asian	37	9.6
European	226	58.5
Hispanic	6	1.5
Indo - Canadian	6	1.5
Middle Eastern	14	3.6
Native Canadian	5	1.3
Unspecified/Unavailable	77	20.1
Diagnosis [range of sperm concentration, mln/mL]		
Fertile pre-vasectomy	37	9.6
Unexplained infertility (15-36 mln/mL)	175 [15-36]	45.3
Oligospermia	174 [0.1-15]	45.1

Table 2. List of down-regulated testis-specific proteins in spermatozoa of rs35033974^{hh} homozygous men, as discovered by shotgun mass spectrometry and measured by SRM.

UniProt accession	Gene name	Protein	Shotgun log2-fold change	SRM log2-fold change	Protein class
Q9BY14	TEX101	Testis-expressed protein 101	-4.8	-4.2	Cell-surface
Q17RY6	LY6K	Lymphocyte antigen 6K	-4.9	-3.8	Cell-surface
P10323	ACR	Acrosin	-2.9	-3.6	Secreted
Q6X784	ZPBP2	Zona pellucida-binding protein 2	-2.9	-1.2	Secreted
Q8IXA5	SPACA3	Sperm acrosome membrane-associated protein 3	-2.6	-2.1	Cell-surface
Q5VZ72	IZUMO3	Izumo sperm-egg fusion protein 3	-2.6	-1.2	Cell-surface
P38567	SPAM1	Hyaluronidase PH-20	-2.6	-3.3	Secreted
Q8NEB7	ACRBP	Acrosin-binding protein	-2.6	-2.9	Secreted
Q9UKF5	ADAM29	Disintegrin and metalloproteinase domain-containing protein 29	-1.3	-1.2	Cell-surface
Q9H4B8	DPEP3	Dipeptidase 3	-0.2	0.4	Cell-surface

FIGURES

Figure 1. Identification of rs35033974 variant in *TEX101* gene. (A) Schematic representation of *TEX101* gene and protein sequence, showing the missense variant c.296 G>T, and the amino acid substitution p.99 G>V. (B) *TEX101* DNA sequencing of WT *TEX101* patient (left panel), rs35033974 heterozygous patient (middle panel) and rs35033974^{hh} homozygous patient (right panel). (C) Scatter plot of sperm concentration and TEX101 levels in seminal plasma of 189 sub-fertile men (diagnosed with unexplained infertility or oligospermia) of European origin. WT, rs35033974 heterozygous and rs35033974^{hh} men are plotted in grey, green and red, respectively. Sperm concentration (15 million/mL) and TEX101 concentration (65 ng/mL) cut-off values are indicated by dotted lines in black. Variant rs35033974 allele frequency is higher for the groups with TEX101<65 ng/mL (58.3% for upper left group and 21% for lower left group), and lower for the groups with TEX101≥65 ng/mL (9.8% for lower right group, and 8.7% for upper right group). (D) Seminal plasma levels of TEX101 normalized by sperm concentration were significantly lower (fold change 4.0, MWU P=0.0005) for G99V heterozygous European men (median 390 ng/mL, N=40), as compared to WT men (median 1,949 ng/mL, N=145). Median values for each group are represented as horizontal lines.

Figure 2. Measurement of WT and G99V variant forms of TEX101 protein by targeted mass spectrometry. (A) Immuno-PRM assay included immunoenrichment of endogenous TEX101

with a mouse monoclonal antibody 34ED556 coupled to sepharose beads, followed by Glu-C digestion and PRM measurements of WT and G99V peptides, as well as a control peptide representing total TEX101. (B) In a WT patient (GG, left panel), high levels of TEX101 were measured in seminal plasma by PRM assay without immunoenrichment. In rs35033974 heterozygote (GT, middle panel) and rs35033974^{hh} homozygote (TT, right panel) patients, WT, G99V variant and control peptides of TEX101 were measured by immuno-PRM assay in spermatozoa lysates. Corresponding heavy isotope-labeled peptides were used as internal standards to ensure correct identification and accurate relative quantification of peaks.

Assuming theoretically equal expression of both alleles and similar ionization efficiencies of WT and G99V peptides, <3% of expected G99V variant form was found in spermatozoa of a heterozygous patient, suggesting that G99V TEX101 form may be degraded during spermatogenesis.

Fig. 3. Identification of proteins down-regulated in rs35033974^{hh} spermatozoa, as compared to WT spermatozoa. (A) Spermatozoa obtained from four WT men with confirmed fertility and four rs35033974^{hh} patients were digested by trypsin, fractionated by strong-cation exchange chromatography, and analyzed by LC-MS/MS. (B) A total of 8,046 protein groups representing 8,473 unique gene names and 8,431 UniProt protein IDs were identified and quantified. Volcano plot revealed proteins down-regulated in rs35033974^{hh} spermatozoa with FDR ≤ 5% (hyperbolic curve) calculated using log₂-transformed fold change ratios, BH-adjusted *t*-test *P*-values and *s*₀=0.4 variance correction. Differentially expressed testis-specific cell-surface and secreted proteins were plotted in blue and green, respectively. Three proteins with adhesion activity of the ADAM family (ADAM18, ADAM29, ADAM32) were identified. Levels of DPEP3, a testis-specific and *TEX101*-interacting protein not affected by *Tex101* knockout in mice, were not changed. No testis-specific cell-surface or secreted proteins were found significantly up-regulated.

Figure 4. Verification of 8 testis-specific cell-surface and secreted proteins in four WT and three rs35033974^{hh} men by SRM and western blotting. (A) A multiplex SRM assay facilitated accurate relative quantification of 8 candidate proteins and confirmed significant down-regulation (>2-fold change, P -value <0.05, dotted lines) of 7 proteins. (B) Western blot analysis of spermatozoa obtained from one WT and one rs35033974^{hh} men confirmed reduced levels of TEX101, LY6K and ADAM29 proteins, while levels of a testis-specific protein DPEP3, present in spermatozoa as both a monomer and a homodimer, were not affected. GAPDH was used as a loading control for total protein.