scientific data

Check for updates

OPEN

DATA DESCRIPTOR

A Comprehensive Proteome of Human Corneal Epithelial Cells Constructed by Cross-platform DIA-Mass Spectrometry

Kenrick Kai-yuen Chan[1,6], Jimmy Ka-wai Cheung[1,6], Shing-yan Roy Chung[1,2], Hang-kin Kong[1,3], Jingfang Bian[2,4], Lei Zhou[1,2,4,5], Chi-wai Do[1,2,4] & Thomas Chuen Lam[1,2,4] ✉

The corneal epithelium serves as the front barrier against environmental stimuli and pathogens on the ocular surface. A comprehensive protein profile of the corneal epithelium would be crucial for understanding the molecular mechanisms that are related to corneal disease. This work demonstrated a library-free data-independent acquisition (DIA) approach across different mass spectrometers and proteomic software to build a comprehensive proteomic dataset for human corneal epithelial cells (HCECs). With the combinational use of different data-independent acquisition technologies of multiple mass spectrometers, including Sciex ZenoTOF 7600 (DIA-SWATH), Bruker TimsTOF Pro2 (DIA-PASEF), and ThermoFisher Orbitrap Fusion Lumos (DIA-HRMS1), protein identification and quantification were performed with superior sensitivity and resolution. By using a library-free DIA approach, this study constructed a more diverse and unbiased proteomic profile of human corneal epithelial cells (HCECs), comprising 11,954 protein groups (1% FDR). This represents the largest corneal proteome reported to date. All raw proteomic data were deposited to ProteomeXchange Consortium via Proteomics Identifications database (PRIDE) with the dataset identifier accession number PXD059451. Our findings hold the potential to enhance future understanding of corneal pathologies and transformative therapeutics.

Background & Summary

Corneal epithelial cells serve as the outermost cellular barrier that plays an important role in maintaining the structural integrity and homeostasis of the ocular surface. Close monitoring of the molecular change in corneal epithelium could provide vital assistance for the management of certain ocular diseases, such as dry eye disease, keratoconus and persistent corneal epithelial dysfunction[1–4]. As proteins regulate a wide range of biological processes, altered proteomic profiles in the cornea could contribute to the mechanistic changes that were induced by specific pathologies. With recent advancements in mass spectrometry, comparing the proteomes of healthy and diseased cornea has become increasingly signified for identifying novel biomarkers and elucidating pathological mechanisms[4–7].

However, significant research gaps exist in the knowledge regarding human corneal epithelial proteomics. This could be attributed to methodological limitations of proteomic technologies and techniques in prior studies as well as the scarcity of human corneal samples. In the field of anterior eye proteomic research, researchers have focused more traditionally on human tear fluid than the cornea due to the non-invasive nature of tear collection, which gives researchers the flexibility in sample collection. In contrast, there was only a limited number of studies that were related to human corneal proteome due to limited available normal

[1]Centre for Eye and Vision Research (CEVR), 17W Hong Kong Science Park, Taipo, Hong Kong. [2]Centre for Myopia Research, School of Optometry, The Hong Kong Polytechnic University, Hung Hom, Hong Kong. [3]Department of Food Science and Nutrition, Faculty of Science, The Hong Kong Polytechnic University, Hung Hom, Hong Kong. [4]Research Centre for SHARP Vision (RCSV), The Hong Kong Polytechnic University, Hung Hom, Hong Kong. [5]Department of Applied Biology and Chemical Technology, Faculty of Science, The Hong Kong Polytechnic University, Hung Hom, Hong Kong. [6]These authors contributed equally: Kenrick Kai-yuen Chan, Jimmy Ka-wai Cheung. ✉e-mail: thomas.c.lam@polyu.edu.hk

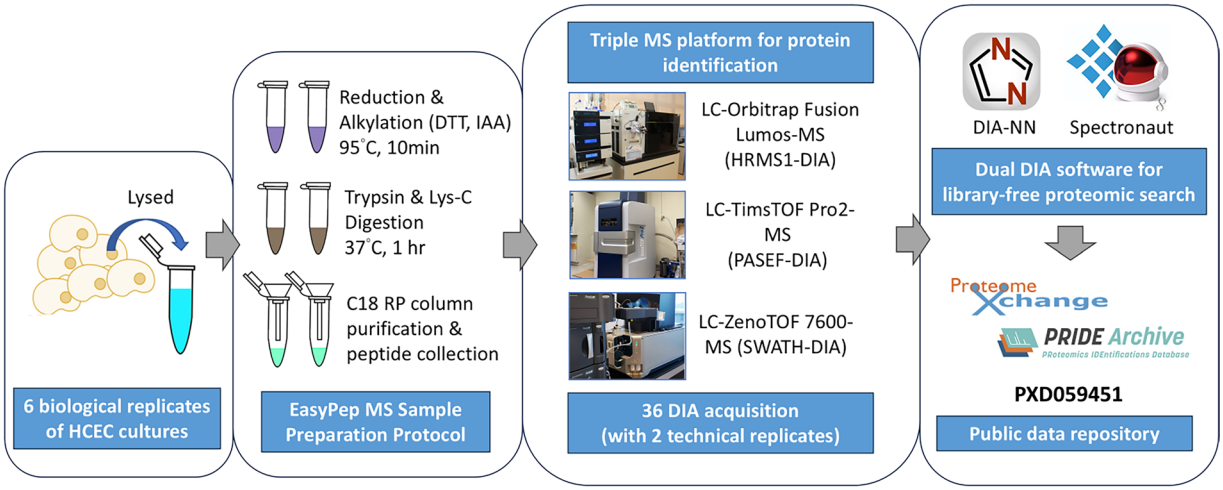

Fig. 1 Experimental design and the workflow for sample preparation.

human corneal samples[8]. Thus, the progress in corneal proteomic research has been lagging far behind and created a remarkable research gap. In terms of biological perspective, studying the proteomic profile of the cornea comprehensively allows the identification of new targets for therapeutic development, including targeted drug delivery as well as drug efficacies. From a technical perspective, most previous proteomic works on human cornea relied heavily on comprehensive spectral libraries generated from Data Dependent Acquisition (DDA) experiments due to limited robustness of older models of mass spectrometers and choice of compatible analysis software, together with more traditional techniques such as gel electrophoresis, resulting in less than 3,000 proteins identified on average[1,5,9–12]. Using a DDA-based approach for proteomic analysis has several downsides that could limit the completeness and accuracy of the results. Due to the bias toward high-abundance peptides of DDA, peptides with low abundance could be overlooked and the difference in methodical acquisition between DDA and Data Independent Acquisition (DIA) could also lead to additional fragmentation variability and discrepancies in peptide identification and quantification[13–15]. Since corneal proteins that might be associated with specific pathologies were often present in low abundance, these candidates that could provide insights into disease mechanisms could be prone to being undetected. Therefore, the reliance of DDA-based approach in prior corneal proteomic study has become another major factor that is attributed to the research gaps. In contrast, without the need for a high-quality preconstructed DDA spectral library, a library-free DIA approach could enhance the reproducibility, dynamic range and flexibility of protein identification[14,16–18].

In response to the growing need for a comprehensive proteomic profile of human cornea epithelial cells (HCECs), this study sought to unveil a more extensive proteome generated by three advanced LC-MS, which were Orbitrap Fusion Lumos (Orbitrap), Trapped Ion Mobility Spectrometry Time-of-Flight Pro2 (TimsTOF) and ZenoTrap Time-of-Flight 7600 (ZenoTOF) Mass Spectrometers, with the use of dual library-free DIA analysis platforms (i.e., Spectronaut and DIA-NN). Each of the abovementioned mass spectrometers utilized distinct DIA technologies that significantly improved their resolution and sensitivity for protein/peptide identification. TimsTOF Pro2 is capable of four-dimensional ion detection that allows the detection of low-abundance proteins in complex biological samples such as whole cell lysate. The design of this instrument could reduce missing values in the data matrix and thus enhance data completeness and robustness of protein identification[19]. Besides, Orbitrap Fusion is recognized for high resolution and sensitivity, dynamic range of MS scan with automatic gain control (AGC)[20]. It is compatible with performing parallel time-constrained processes that maximize the efficiency during data acquisition[21]. Meanwhile, the unique Zeno trap technology of ZenoTOF 7600 allows longer ion accumulation times. This also helps to improve the detection of low-abundance proteins. The coupled microflow LC setup and high-speed acquisition capabilities of the system facilitate rapid identification and quantitation of proteins[22]. Apart from the mass spectrometers, Spectronaut and DIA-NN are both recent proteomic tools that are equipped with machine learning techniques to further improve the accuracy and efficiency of peptide/protein identification and hence facilitate effective classification of useful biomarkers and pathological pathways[23–25].

The implementation of advanced multiple LC-MS platforms coupled with cutting-edge proteomic bioinformatics facilitated the generation of a comprehensive proteomic profile of HCECs in this study. Here we present the most comprehensive combined proteome containing 11,954 protein groups at 1% FDR. This proteomic dataset was generated by combining the results of DIA from a total of 36 injections, 12 injections performed by each of the TimsTOF, ZenoTOF and Orbitrap MS, with peptide samples prepared from 6 biological replicates of HCECs culture and with 2 technical replicates of each injection (Fig. 1). As this proteome was built on the basis of immortalized HCECs that were cultured in optimal and unstressed conditions, it has the potential to become an important reference for *in vitro* studies of proteomic alteration in corneal pathologies.

Methods

Cell culture. Immortalized HCECs (#BNCC337876, BeNa Culture Collection, China) were cultured with Dulbecco's Modified Eagle Medium F12 (DMEM/F12; #11320033, ThermoFisher, USA) together with 10% fetal bovine serum (FBS; #A5256701, ThermoFisher, USA) and 1% penicillin-streptomycin antibiotic mixture (P/S; #15140122, ThermoFisher, USA) at standard culture conditions[26]. To prevent cell differentiation from cellular stress, culture medium was renewed every 48 hours and passage of cell culture was performed whenever the culture reached 80% confluence. HCECs at passage 6 were used in this study. The medium of HCECs was changed to FBS-free DMEM/F12 with 1% P/S and then incubated at standard culture conditions for 24 hours prior to protein extraction.

Protein extraction and sample preparation for mass spectrometry. HCECs were washed with ice-cold 1X phosphate buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) six times. EasyPep Lysis buffer was then added to lyse the HCECs. The concentration of the proteins extracted was further adjusted for subsequent sample digestion and purification using EasyPep MS sample preparation kit (#A40006, ThermoFisher, USA) with the manufacturer's recommended protocol. In brief, proteins were reduced and alkylated simultaneously with reduction and alkylation solution and incubated at 95 °C for 10 mins. Fifty micrograms of proteins from each individual sample were then digested with Trypsin and Lys-C at 37 °C for 1 hour. Digested peptides were cleaned up with the aid of reverse phase-based C18 spin column and wash solutions. Purified peptides were dried and then reconstituted with 0.1% formic acid (FA). Peptides were quantified using peptide assay (#23275, ThermoFisher, USA) and diluted to the concentration of 0.25 µg/µL prior to mass spectrometric analysis, with an equal loading amount of 1 µg injection.

Liquid Chromatography-Mass Spectrometry (LC-MS). One microgram of the purified peptide samples was analyzed by LC-coupled Orbitrap Fusion Lumos Mass Spectrometer (Orbitrap; ThermoFisher, USA), Nano LC-coupled Trapped Ion Mobility Spectrometry Time-of-Flight Pro2 Mass Spectrometer (TimsTOF; Bruker, USA) and Micro LC-coupled ZenoTrap Time-of-Flight 7600 Mass Spectrometer (ZenoTOF; SCIEX, USA) in data independent acquisition (DIA) mode respectively. For the sample fractionation in all three LC-MS, solvent A (0.1% FA in H_2O) and solvent B (0.1% FA in acetonitrile) were used as the gradient. Optimized MS settings were applied to maximize the performance on protein identifications.

For Orbitrap, peptides were loaded onto the Acclaim™ PepMap™ 100 HPLC Column (100 Å, 5 µm, 5 mm × 1 mm, C18, Thermo Scientific, USA) by loading buffer (0.1% formic acid in water) at 0.3 µL/min for 2 mins, and then separated on Aurora Ultimate nanoflow UHPLC column (120 Å, 1.7 µm, 25 cm × 75 µm; Ionopticks, Australia) using Dionex UltiMate 3000 RSLCnano NanoLC system (Thermo Scientific, USA) connect to the Orbitrap Fusion. The peptides were then fractionated by a 90-minute linear elution gradient (0–2 min: 2–6%B, 2–79 min: 6–30% B, 79–82 min: 30–90%B, 82–87 min: 90%B, 87–90 min: 2%B). For the High-Resolution MS1 Data-Independent Acquisition (HRMS1-DIA) experiment, the MS1 scan range was set at 400–1500 m/z in positive mode with maximal injection time of 50 ms, orbitrap resolution at 30,000 and standard AGC target. While for the data independent analysis, scan range was set 200–2000 m/z with orbitrap resolution at 7,500 and HCD collision energy and RF lens were set at 30%. Ten 15 m/z isolation windows and twenty-five 10 m/z isolation windows were assigned within the mass range of 400–650 m/z and 650–800 m/z respectively with 1 m/z of window overlap in quadrupole mode.

For TimsTOF, peptides were loaded onto the Acclaim™ PepMap™ 100 HPLC Column by loading buffer (0.1% FA in H_2O) at 0.3 µL/min for 2 mins, and then separated on an Aurora Ultimate CSI Column (120 Å, 25 cm × 75 µm, 1.7 µm, C18, Ionopticks, AU) using Dionex UltiMate 3000 RSLCnano NanoLC system connected to the TimsTOF Pro2. The peptides were then fractionated by a 45-minute linear elution gradient (0–2 min: 2%B, 2–2.1 min: 2–5%B, 2.1–33 min: 5–35%B, 33–33.5 min: 35–90%B, 33.5–37 min: 90%B, 37–37.1 min: 90-2%B, 37.1–45 min: 2%B). For Data-Independent Acquisition with Parallel Accumulation Serial Fragmentation (DIA-PASEF) experiments, the range of MS1 scan was 100-1,700 m/z in positive mode. Twenty-one 25 m/z isolation windows were assigned over the mass range of 475–1,000 m/z with 100% duty cycle, 100 ms accumulation and ramp time in the TIMS.

For ZenoTOF, peptides were loaded on the nanoEase M/Z Symmetry trap column (100 Å, 300 µm x 50 mm, 5 µm C18, Waters, USA) by loading buffer (0.1% FA in H_2O) at 10 µL/min for 3 mins, and then separated on a nanoEase M/Z HSS T3 Column (1.8 µm, 75 µm x 200 mm, C18, Waters, USA) using an MicroLC system (UPLC M-Class, Waters) connected to ZenoTOF. The peptides were then fractionated by a 15-minute linear elution gradient (0-1 min: 3%B, 1–9 min: 10%B, 9–19 min: 30%B, 10–10.2 min: 40%B, 10.2–12.1 min: 95%B, 12.1–15 min: 3%B). For Data-Independent Acquisition with sequential window acquisition of all theoretical fragment ion spectra (DIA-SWATH) experiments, isolation width from 3–100 Da is set in a looped mode over the full mass range (100–1,500 m/z) scan, and 64 overlapping windows were constructed, with 20 ms accumulation time per cycle using Zeno-pulsing.

Proteomic analysis. The raw data files generated from ZenoTOF, Orbitrap and TimsTOF were imported to Spectronaut (V.19; Biognosys, Switzerland) and DIA-NN (V 1.9.1; Cambridge, UK) for identifying and quantifying the relative abundances of the proteins in the samples. For Spectronaut, proteomic analysis was conducted in library-free directDIA mode. Reviewed *Homo sapiens* protein sequence database (FASTA) from UniProt and gene ontology annotation (GOA) of *Homo sapiens* from Gene Ontology (retrieved on 27-Dec-2023) were imported for proteomic search and functional annotation. The FASTA file consisted of 20,428 protein sequences including isoform and the GOA composed of 19,653 entries. Proteomic identification and quantification were run with the BGS factory default settings with cleavage rules of Trypsin/P and LysC/P applied. For DIA-NN, library-free search and library generation were conducted with the same database (FASTA) mentioned above, with deep

	Spectronaut			DIA-NN		
	TimsTOF	ZenoTOF	Orbitrap	TimsTOF	ZenoTOF	Orbitrap
Average No. of precursor	76,733	48,140	61,182	55,568	46,054	57,038
Median CV (%) at precursor level	22.5	21.5	13.2	—	—	—
Average No. of protein group	8,223	4,740	5,935	7,300	4,929	5,800
Median CV (%) at protein group level	—	—	—	11.5	5.6	5.8
Average missed cleavage	0.25	0.14	0.17	0.19	0.11	0.14
Average No. of precursor per protein group	9.3	10.2	10.3	7.6	9.3	9.8

Table 1. Analysis overview of the proteomic dataset generated from individual LC-MS and software.

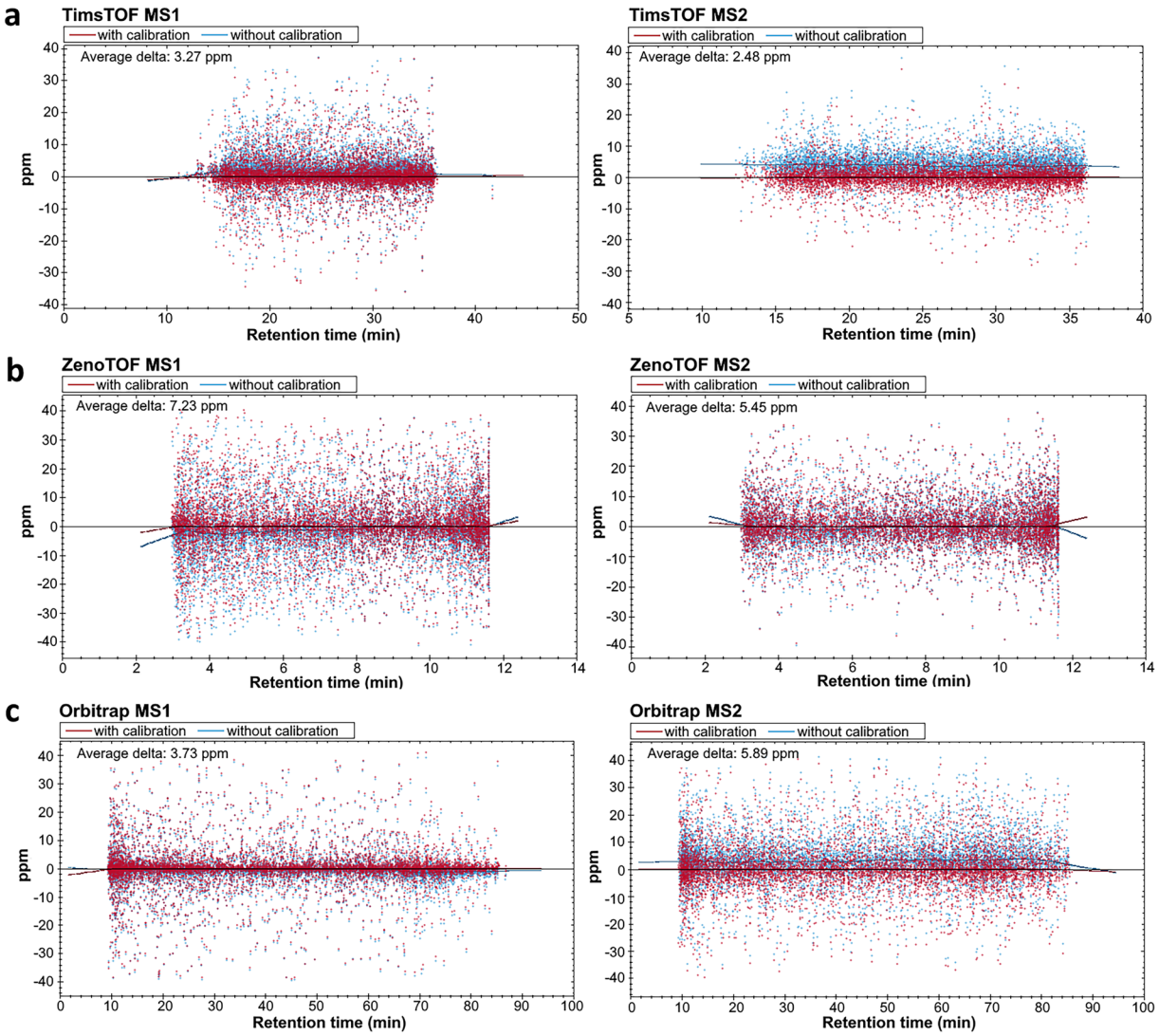

Fig. 2 The average mass error observed in MS1 and MS2 scans of the mass spectrometers. Blue and red dots represented the original and calibrated mass errors that were recorded during the acquisitions respectively. **(a)** In TimsTOF, high mass accuracy was observed with an average mass error of 3.27 ppm in MS1 scan and 2.48 ppm in MS2 scan. **(b)** In ZenoTOF, an acceptable mass accuracy was observed with an average mass error of 7.23 ppm in MS1 scan and 5.45 ppm in MS2 scan. **(c)** In Orbitrap, high mass accuracy was observed with an average mass error of 3.73 ppm in MS1 scan and 5.89 ppm in MS2 scan.

learning-based spectra, RTs and IMs prediction enabled and cleavage rule of Trypsin/P applied. All proteomic searches were conducted with 1% FDR. The resultant proteomic profiles that were generated from DIA-NN and Spectronaut were integrated on the basis of protein group identity. The consistency of the protein groups was assessed by calculating the percentage of successful identifications relative to the total number of biological and technical replicates performed.

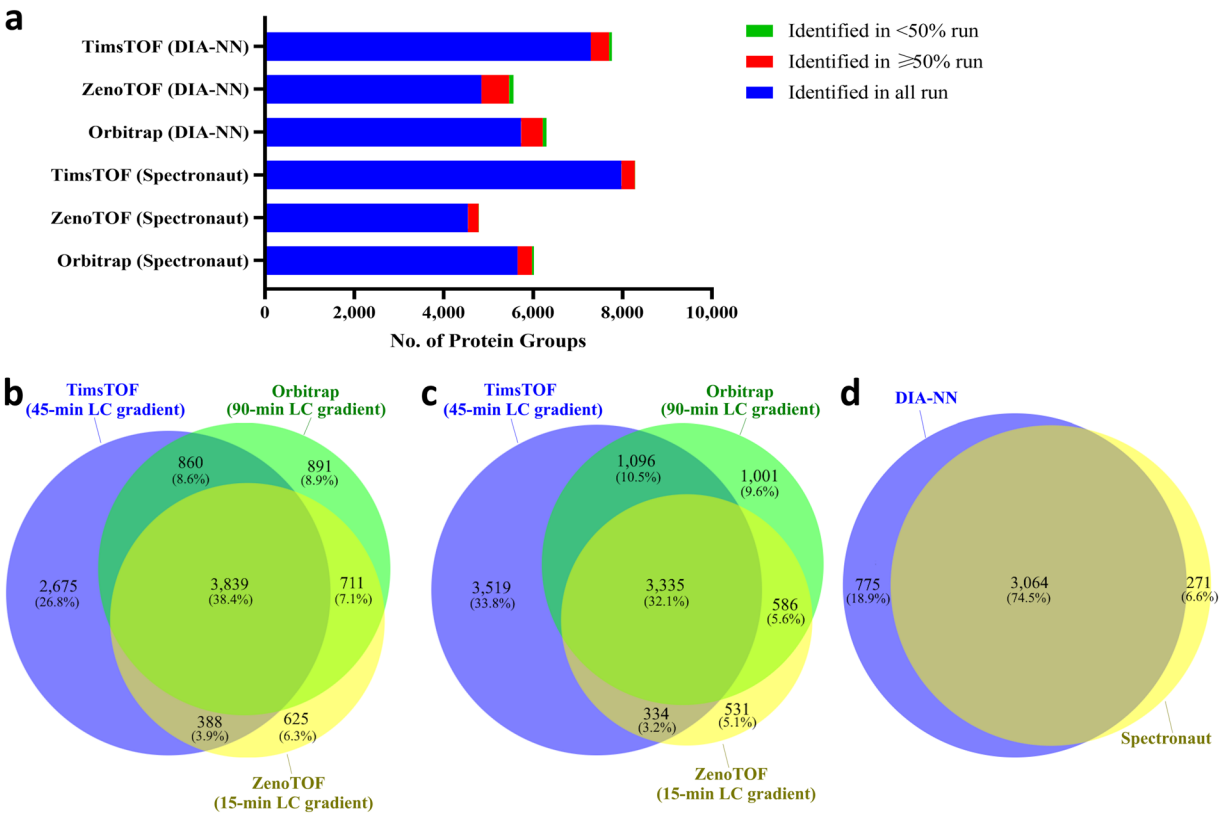

Fig. 3 Consistency of the identification of protein group. **(a)** Among all runs with different combinations of MS (TimsTOF with 45-min LC gradient, ZenoTOF with 15-min LC gradient and Orbitrap with 90-min LC gradient) and proteomic software (DIA-NN and Spectronaut), the consistency of the identification of protein group were categorized into three classes accordingly. A total of 38,714 protein group identification were performed in these listed combinations, 93.1% of the identification were found in all replicates (blue), 6.1% of the identification were found in more than half of the replicates (red) and 0.8% of the identification were found in less than half of the replicates (green). **(b)** Within the dataset generated by DIA-NN, a total of 9,989 protein groups were profiled. Among the subsets of data generated by TimsTOF, ZenoTOF and Orbitrap, 3,839 (38.4%) protein groups were commonly identified. **(c)** Within the dataset generated by Spectronaut, a total of 10,402 protein groups were identified and 3,335 (32.1%) of these protein groups were commonly found across the three subsets of data from the respective MS. **(d)** When overlapping the common protein groups from the DIA-NN (intersection of Fig. 3b) and Spectronaut datasets (intersection of Fig. 3c), 3,064 (74.5%) of these protein groups were able to be identified by both software.

Data Records

All raw proteomic data of this study were deposited to ProteomeXchange Consortium via Proteomics Identifications database (PRIDE) with the dataset identifier accession number PXD059451[27]. For data acquired by TimsTOF Pro2, the whole folder for each individual sample is in .d format and zipped, with each naming with the number following the letter R for biological replicates and the number following the letter T for technical replicates. For data acquired from Orbitrap Fusion, each uploaded raw file consisted of 2 native file types known as .raw and .raw.quant. For each raw file, the naming with the number following the letter R for biological replicates and the number following the letter T for technical replicates. For data acquired by ZenoTOF 7600, each uploaded raw file consisted of 3 native file types known as .wiff, .wiff2 and .wiff.quant. with each naming with the number following the letter N for biological replicates and the number at the end for technical replicates. There are three .tsv file-based outputs generated from the software DIA-NN, and three .sne file-based outputs generated from the software Spectronaunt. In this work, a total list of protein groups that were identified from HCECs was generated and provided in the repository as an .xslx file titled "SupplementaryTable 1_protein_List_integrity_and_consistency".

Technical Validation

Coverage, integrity and consistency of proteome. To ensure the integrity and consistency of the proteome, the dataset that was generated by each individual LC-MS was examined using Spectronaut and DIA-NN. The average missed cleavage, data completeness, coefficient of variation (CV) and mass accuracy of all runs in this work lay within the acceptable range[18,28,29] (Table 1 & Fig. 2). In terms of the consistency of protein group identification within an individual MS, 93.1% of the total identification was consistently found in all biological and

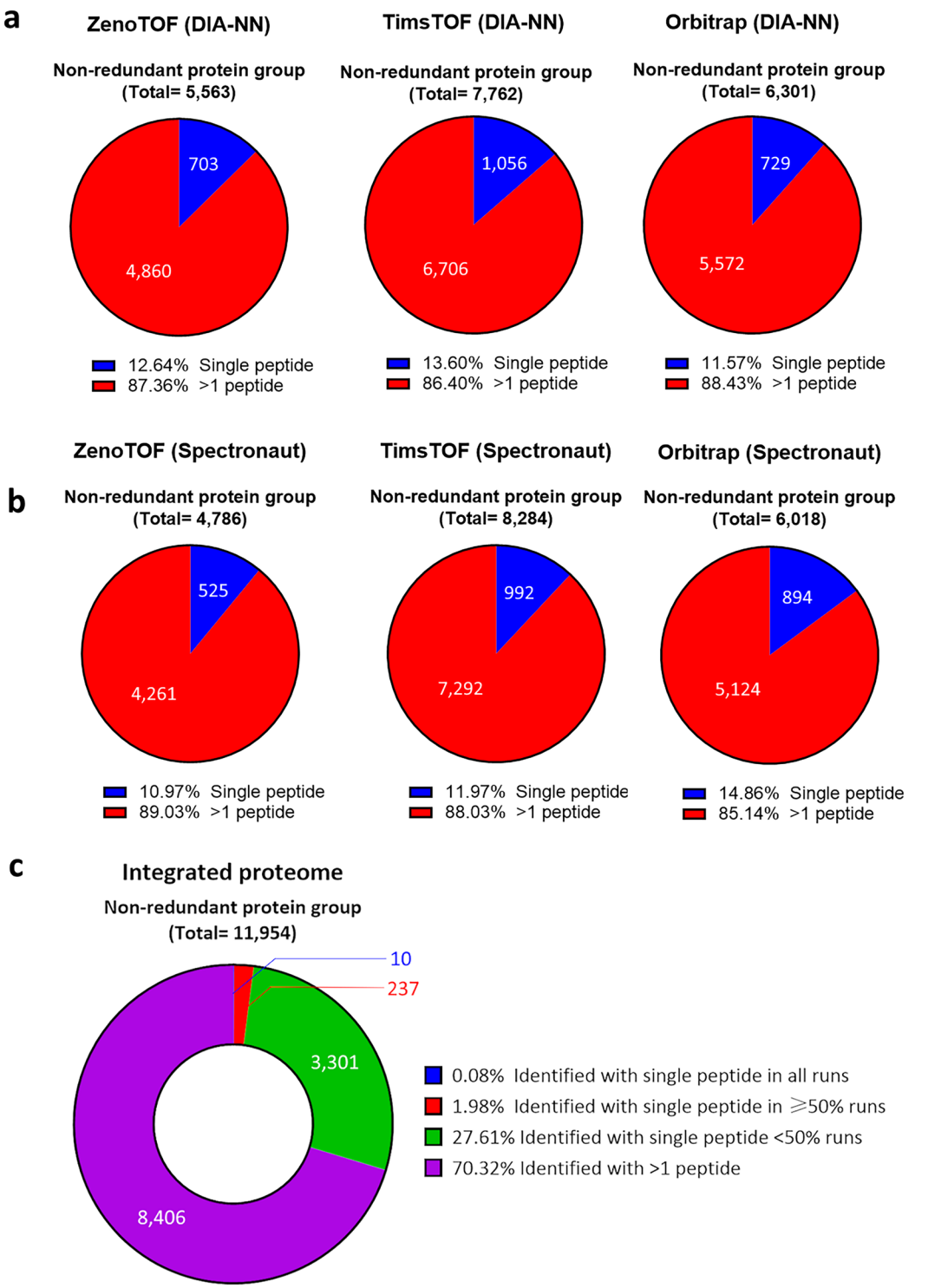

Fig. 4 The ratio of protein group identified with single peptide. **(a)** In the data subset processed with DIA-NN, 12.6% of the protein groups identified by ZenoTOF, 13.6% of the protein groups identified by TimsTOF, and 11.6% of the protein groups identified by Orbitrap were single peptide-containing protein groups. The average percentage of protein groups identified with single peptide by the three MS was 12.6%. **(b)** While in the data subset processed with Specotronaut, 11.0% of the protein groups identified by ZenoTOF, 12.0% of the protein groups identified by TimsTOF, and 14.9% of the protein groups identified by Orbitrap were single peptide-containing protein groups. The average percentage of protein groups identified with single peptide by the three MS was also 12.6%. **(c)** After the integration of the datasets, 70.3% of the protein groups were identified with more than one peptide and 0.08% of the protein groups were consistently identified with single peptide among all combinations of MS and analysis software. While 1.98% and 27.6% of the protein groups were identified with single peptide among ≥50% and <50% of the combinations of MS and analysis software respectively.

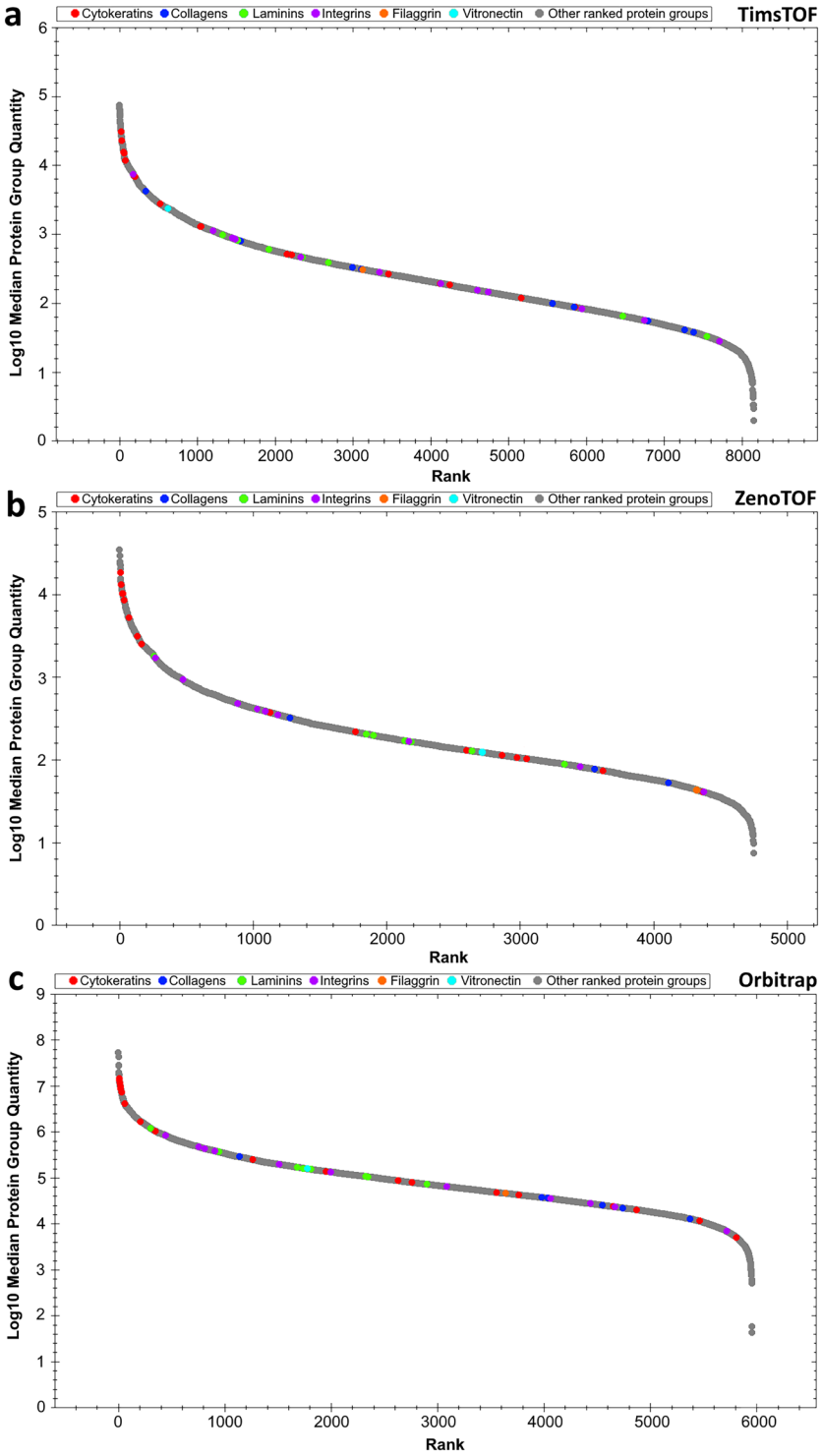

Fig. 5 The ranking and distribution of signature proteins among all ranked protein groups. **(a)** In the dataset of TimsTOF, 17 types of cytokeratins (keratin 1, 2, 5, 6 A, 7–10, 13–19, 74 & 80), 12 types of integrins (integrin α1-6, αV, α9, β1 & β4-6), 7 types of laminins (laminin subunit α3, α5, β1-3 & γ1-2), 7 types of collagens (collagen type 4, 7, 11-12 & 16–18) vitronectin and filaggrin were identified. **(b)** In the dataset of ZenoTOF, 16 types of cytokeratins (keratin 1, 5, 6 A, 7–10, 13–20 & 80), 10 types of integrins (integrin α2-4, αV, α6, β1 & β4-6), 6 types of laminins (laminin subunit α3, α5, β1, β3 & γ1-2), 2 types of collagens (collagen type 4 & 17), vitronectin and filaggrin were identified. **(c)** In the dataset of Orbitrap, 20 types of cytokeratins (keratin 1, 2, 4, 5, 6 A, 7–10, 12–19, 28, 72 & 80), 11 types of integrins (integrin α2-6, αV, β1, β4-6, β8), 6 types of laminins (laminin subunit α3, α5, β1-3 & γ1-2), 5 types of collagens (collagen type 4, 11-12, 17, 22), vitronectin and filaggrin were identified.

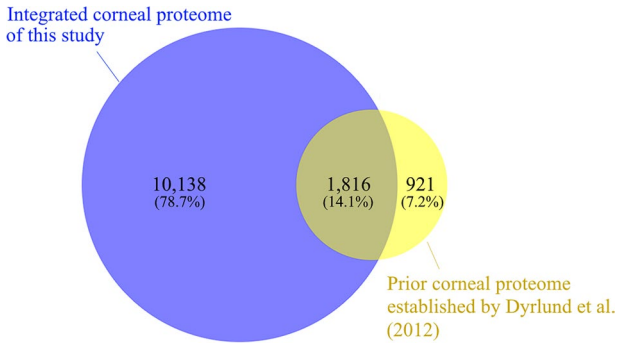

Fig. 6 Overlapping between the integrated corneal proteome of this study and the previous largest corneal proteome reported[12]. This integrated dataset reproduced 1,816 proteins that were reported in the dataset of Dyrlund *et al*. The integrated dataset consisted of 10,138 proteins that were not found in the prior dataset.

technical replicates (Fig. 3). Together with the low CV (<12%) observed at protein group level (Table 1), these indicated the high consistency of the identification process.

Moreover, comparison of the proteome was extended across different MS and software to further assess the consistency and integrity of these protein groups on a cross-platform basis (Fig. 3). We would like to stress that 24.2% of the total non-redundant protein groups were core protein groups, which were identified with 100% consistency[27]. Compared to the proteome identified with individual MS and software, the total number of the protein groups was increased by 185%. Also, an average of 12.6% of the protein groups were identified with single peptide-containing by a respective single MS and analysis software combination (Fig. 4a,b). By validating the profile of single peptide containing protein group across the subset, only 0.08% of the protein groups in the integrated proteome were consistently identified with single peptide among all combinations of MS and software (Fig. 4c)[27]. In other words, cross-platform strategy could enhance the integrity of protein identification by complementing the profile of single peptide containing protein group across the subset.

In terms of coverage of the proteome, our integrated proteomic dataset covered a wide range of signature structural proteins of human corneal epithelia that have been reported previously[30–42], which included 28 types of cytokeratins, 9 types of laminins, 22 types of integrins, 15 types of collagens, filaggrin and vitronectin (Fig. 5). These matches validated the integrity of our proteomic dataset. Compared to the prior largest proteome of human corneal epithelium[12], the number of proteins identified in our integrated proteome was four times larger (Fig. 6). Also, the integrated proteome managed to reproduce 66.4% of the prior proteome and provided 10,138 extra proteins that were not found in the prior dataset. Our dataset represents the largest corneal proteome reported to date.

Code availability

This study did not use any custom computer code or algorithm to generate the data.

Received: 13 November 2024; Accepted: 14 April 2025;
Published online: 23 May 2025

References

1. Yam, G. H.-F. *et al*. Differential epithelial and stromal protein profiles in cone and non-cone regions of keratoconus corneas. *Scientific Reports* **9**, 2965, https://doi.org/10.1038/s41598-019-39182-6 (2019).
2. Cai, J., Estes, A. & Liu, Y. Omics Analyses in Keratoconus: from Transcriptomics to Proteomics. *Current Ophthalmology Reports* **8**, 216–225, https://doi.org/10.1007/s40135-020-00253-x (2020).
3. Lin, J. B. *et al*. Dry eye disease in mice activates adaptive corneal epithelial regeneration distinct from constitutive renewal in homeostasis. *Proc Natl Acad Sci USA* **120**, e2204134120, https://doi.org/10.1073/pnas.2204134120 (2023).
4. Moon, C. E. *et al*. Integrated Analysis of Transcriptome and Proteome of the Human Cornea and Aqueous Humor Reveal Novel Biomarkers for Corneal Endothelial Cell Dysfunction. *Int J Mol Sci* **24** https://doi.org/10.3390/ijms242015354 (2023).
5. Kang, B. S., Lam, T. C., Cheung, J. K., Li, K. K. & Kee, C. S. Corneal proteome and differentially expressed corneal proteins in highly myopic chicks using a label-free SWATH-MS quantification approach. *Sci Rep* **11**, 5495, https://doi.org/10.1038/s41598-021-84904-4 (2021).
6. Chaerkady, R. *et al*. The keratoconus corneal proteome: loss of epithelial integrity and stromal degeneration. *Journal of proteomics* **87**, 122–131 (2013).
7. Nielsen, K. *et al*. Proteome profiling of corneal epithelium and identification of marker proteins for keratoconus, a pilot study. *Experimental eye research* **82**, 201–209 (2006).
8. Wu, P.-S. *et al*. Comparative Proteomics Reveals Prolonged Corneal Preservation Impaired Ocular Surface Immunity Accompanied by Fibrosis in Human Stroma. *Journal of Proteome Research* **22**, 3730–3741, https://doi.org/10.1021/acs.jproteome.3c00383 (2023).
9. Galiacy, S. D. *et al*. Deeper in the human cornea proteome using nanoLC–Orbitrap MS/MS: An improvement for future studies on cornea homeostasis and pathophysiology. *Journal of Proteomics* **75**, 81–92, https://doi.org/10.1016/j.jprot.2011.09.020 (2011).
10. Joseph, R., Srivastava, O. P. & Pfister, R. R. Differential epithelial and stromal protein profiles in keratoconus and normal human corneas. *Experimental Eye Research* **92**, 282–298, https://doi.org/10.1016/j.exer.2011.01.008 (2011).
11. Deng, M. *et al*. Proteomic profiling of human corneal stroma from long-term contact lens wearers reveals activation of inflammatory responses. *Contact Lens and Anterior Eye* **46**, 101820, https://doi.org/10.1016/j.clae.2023.101820 (2023).
12. Dyrlund, T. F. *et al*. Human cornea proteome: identification and quantitation of the proteins of the three main layers including epithelium, stroma, and endothelium. *J Proteome Res* **11**, 4231–4239, https://doi.org/10.1021/pr300358k (2012).

13. Searle, B. C. *et al*. Generating high quality libraries for DIA MS with empirically corrected peptide predictions. *Nature Communications* **11**, 1548, https://doi.org/10.1038/s41467-020-15346-1 (2020).
14. Pino, L. K., Just, S. C., MacCoss, M. J. & Searle, B. C. Acquiring and Analyzing Data Independent Acquisition Proteomics Experiments without Spectrum Libraries. *Mol Cell Proteomics* **19**, 1088–1103, https://doi.org/10.1074/mcp.P119.001913 (2020).
15. Heil, L. R. *et al*. Building Spectral Libraries from Narrow-Window Data-Independent Acquisition Mass Spectrometry Data. *J Proteome Res* **21**, 1382–1391, https://doi.org/10.1021/acs.jproteome.1c00895 (2022).
16. Zhang, F. *et al*. A Comparative Analysis of Data Analysis Tools for Data-Independent Acquisition Mass Spectrometry. *Molecular & Cellular Proteomics* **22** https://doi.org/10.1016/j.mcpro.2023.100623 (2023).
17. Sinitcyn, P. *et al*. MaxDIA enables library-based and library-free data-independent acquisition proteomics. *Nature Biotechnology* **39**, 1563–1573, https://doi.org/10.1038/s41587-021-00968-7 (2021).
18. Li, J., Smith, L. S. & Zhu, H. J. Data-independent acquisition (DIA): An emerging proteomics technology for analysis of drug-metabolizing enzymes and transporters. *Drug Discov Today Technol* **39**, 49–56, https://doi.org/10.1016/j.ddtec.2021.06.006 (2021).
19. Lou, R. *et al*. Benchmarking commonly used software suites and analysis workflows for DIA proteomics and phosphoproteomics. *Nature Communications* **14**, 94, https://doi.org/10.1038/s41467-022-35740-1 (2023).
20. Greguš, M., Koller, A., Ray, S. & Ivanov, A. R. Improved Data Acquisition Settings on Q Exactive HF-X and Fusion Lumos Tribrid Orbitrap-Based Mass Spectrometers for Proteomic Analysis of Limited Samples. *Journal of Proteome Research* **23**, 2230–2240, https://doi.org/10.1021/acs.jproteome.4c00181 (2024).
21. Ludwig, C. *et al*. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol Syst Biol* **14**, e8126, https://doi.org/10.15252/msb.20178126 (2018).
22. Wang, Z. *et al*. High-throughput proteomics of nanogram-scale samples with Zeno SWATH MS. *Elife* **11** https://doi.org/10.7554/eLife.83947 (2022).
23. Swan, A. L., Mobasheri, A., Allaway, D., Liddell, S. & Bacardit, J. Application of machine learning to proteomics data: classification and biomarker identification in postgenomics biology. *Omics* **17**, 595–610, https://doi.org/10.1089/omi.2013.0017 (2013).
24. Vizza, P. *et al*. Machine learning pipeline to analyze clinical and proteomics data: experiences on a prostate cancer case. *BMC Medical Informatics and Decision Making* **24**, 93, https://doi.org/10.1186/s12911-024-02491-6 (2024).
25. Hartman, E. *et al*. Interpreting biologically informed neural networks for enhanced proteomic biomarker discovery and pathway analysis. *Nature Communications* **14**, 5359, https://doi.org/10.1038/s41467-023-41146-4 (2023).
26. Wong, K. Y., Liu, Y., Wong, M. S. & Liu, J. Cornea-SELEX for aptamers targeting the surface of eyes and liposomal drug delivery. *Exploration (Beijing)* **4**, 20230008, https://doi.org/10.1002/exp.20230008 (2024).
27. Lam, T. C. & Cheung, J. K. *PRIDE* https://identifiers.org/pride.project:PXD059451 (2025).
28. Fischer, F. & Poetsch, A. Protein cleavage strategies for an improved analysis of the membrane proteome. *Proteome Sci* **4**, 2, https://doi.org/10.1186/1477-5956-4-2 (2006).
29. Gillet, L. C., Leitner, A. & Aebersold, R. Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing. *Annual Review of Analytical Chemistry* **9**, 449–472, https://doi.org/10.1146/annurev-anchem-071015-041535 (2016).
30. Ligocki, A. J. *et al*. Molecular characteristics and spatial distribution of adult human corneal cell subtypes. *Scientific Reports* **11**, 16323, https://doi.org/10.1038/s41598-021-94933-8 (2021).
31. Yeung, V. *et al*. Proteomic Characterization of Corneal Epithelial and Stromal Cell-Derived Extracellular Vesicles. *International Journal of Molecular Sciences* **25**, 10338 (2024).
32. Kitazawa, K. *et al*. Gene expression signatures of human senescent corneal and conjunctival epithelial cells. *Aging (Albany NY)* **15**, 9238–9249, https://doi.org/10.18632/aging.205113 (2023).
33. Ortiz-Melo, M. T., Garcia-Murillo, M. J., Salazar-Rojas, V. M., Campos, J. E. & Castro-Muñozledo, F. Transcriptional profiles along cell programming into corneal epithelial differentiation. *Exp Eye Res* **202**, 108302, https://doi.org/10.1016/j.exer.2020.108302 (2021).
34. Tong, L. *et al*. Expression and Regulation of Cornified Envelope Proteins in Human Corneal Epithelium. *Investigative Ophthalmology & Visual Science* **47**, 1938–1946, https://doi.org/10.1167/iovs.05-1129 (2006).
35. Byström, B., Virtanen, I., Rousselle, P., Gullberg, D. & Pedrosa-Domellöf, F. Distribution of Laminins in the Developing Human Eye. *Investigative Ophthalmology & Visual Science* **47**, 777–785, https://doi.org/10.1167/iovs.05-0367 (2006).
36. Szaflik, J. P. *et al*. Genetics of Meesmann corneal dystrophy: a novel mutation in the keratin 3 gene in an asymptomatic family suggests genotype-phenotype correlation. *Molecular Vision* **14**, 1713 (2008).
37. Zimmermann, D. R., Trüeb, B., Winterhalter, K. H., Witmer, R. & Fischer, R. W. Type VI collagen is a major component of the human cornea. *FEBS letters* **197**, 55–58 (1986).
38. Cameron, J. D., Skubitz, A. P. & Furcht, L. T. Type IV collagen and corneal epithelial adhesion and migration. Effects of type IV collagen fragments and synthetic peptides on rabbit corneal epithelial cell adhesion and migration *in vitro*. *Investigative Ophthalmology & Visual Science* **32**, 2766–2773 (1991).
39. McKay, T. B., Schlötzer-Schrehardt, U., Pal-Ghosh, S. & Stepp, M. A. Integrin: Basement membrane adhesion by corneal epithelial and endothelial cells. *Experimental eye research* **198**, 108138 (2020).
40. Stepp, M. A. Corneal integrins and their functions. *Experimental Eye Research* **83**, 3–15, https://doi.org/10.1016/j.exer.2006.01.010 (2006).
41. Chow, S. & Di Girolamo, N. Vitronectin: a migration and wound healing factor for human corneal epithelial cells. *Invest Ophthalmol Vis Sci* **55**, 6590–6600, https://doi.org/10.1167/iovs.14-15054 (2014).
42. Ainscough, S. L., Barnard, Z., Upton, Z. & Harkin, D. G. Vitronectin supports migratory responses of corneal epithelial cells to substrate bound IGF-I and HGF, and facilitates serum-free cultivation. *Exp Eye Res* **83**, 1505–1514, https://doi.org/10.1016/j.exer.2006.08.012 (2006).

Acknowledgements

This work was financially supported by the InnoHK initiative and the Hong Kong Special Administrative Region Government and the Research Centre for SHARP Vision (P0043871, P0039545). The authors gratefully acknowledge technical support from the University Research Facility in Chemical and Environmental Analysis (UCEA) and the University Research Facility in Life Sciences (ULS) of The Hong Kong Polytechnic University.

Author contributions

K.K.Y. Chan, J.K.W. Cheung and T.C. Lam conceived the project. K.K.Y. Chan, J.K.W. Cheung drafted the manuscript and performed the experiments and data analysis. S.Y.R. Chung assisted in the experiments and sample preparations. T.C. Lam, L. Zhou, C.W. Do, J. Bian and H.K. Kong provided intellectual input and fund support. All authors reviewed and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to T.C.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.