

New Results

Follow this preprint Previous

Next

Protein sequencing with single amino acid resolution discerns peptides that discriminate tropomyosin proteoforms

Natchanon Sittipongpittaya, Kenneth Skinner, Erin Jeffery, Emily Watts-Whitehead, Gloria M Sheynkman

doi: <https://doi.org/10.1101/2024.11.04.621980>

This article is a preprint and has not been certified by peer review [what does this mean?].



Abstract Info/History Metrics Preview PDF

Abstract

Protein variants of the same gene--proteoforms--can have high molecular similarity yet exhibit different biological functions. Thus, identifying unique peptides that unambiguously map to proteoforms can provide crucial biological insights. In humans, four human tropomyosin (TPM) genes produce similar proteoforms that can be challenging to distinguish with standard proteomics tools. For example, TPM1 and TPM2 share 85% sequence identity, with amino acid substitutions that play unique roles in muscle contraction and myopathies. In this study, we evaluated the ability of the recently released Platinum single-molecule protein sequencer to detect proteoform-informative peptides. Platinum employs fluorophore-labeled recognizers that reversibly bind to cognate N-terminal amino acids (NAAs), enabling polypeptide sequencing within nanoscale apertures of a semiconductor chip that can accommodate single peptide molecules. As a proof of concept, we evaluated the ability of Platinum to distinguish three main types of proteoform variation: paralog-level, transcript-level, and post-translational modification (PTM). We distinguished paralogous TPM1 and TPM2 peptides differing by a single isobaric residue (leucine/isoleucine). We also distinguished tissue-specific TPM2 spliceforms. Notably, we found that a phosphotyrosine-modified peptide displayed reduced recognizer affinity for tyrosine, showing sensitivity to PTMs. This study paves the way for the targeted detection of proteoform biomarkers at the single molecule level

Competing Interest Statement

KAS is an employee and shareholder of Quantum-Si. No other authors have any disclosures.

Copyright The copyright holder for this preprint is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

bioRxiv and medRxiv thank the following for their generous financial support:

The Chan Zuckerberg Initiative, Cold Spring Harbor Laboratory, the Sergey Brin Family Foundation, California Institute of Technology, Centre National de la Recherche Scientifique, Fred Hutchinson Cancer Center, Imperial College London, Massachusetts Institute of Technology, Stanford University, University of Washington, and Vrije Universiteit Amsterdam.

Back to top

Posted November 06, 2024.

Download PDF
Print/Save Options

Email
Share
Citation Tools
Get QR code

Post Like 0

COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv

Subject Area

Genomics

Subject Areas

All Articles

Animal Behavior and Cognition
Biochemistry
Bioengineering
Bioinformatics
Biophysics
Cancer Biology
Cell Biology
Clinical Trials*
Developmental Biology
Ecology
Epidemiology*
Evolutionary Biology
Genetics
Genomics
Immunology
Microbiology
Molecular Biology
Neuroscience
Paleontology
Pathology
Pharmacology and Toxicology
Physiology
Plant Biology
Scientific Communication and Education
Synthetic Biology
Systems Biology
Zoology

* The Clinical Trials and Epidemiology subject categories are now closed to new submissions following the completion of bioRxiv's clinical research pilot project and launch of the dedicated health sciences server medRxiv (submit.medrxiv.org). New papers that report results of Clinical Trials must now be submitted to medRxiv. Most new Epidemiology papers also should be submitted to medRxiv, but if a paper contains no health-related information, authors may choose to submit it to another bioRxiv subject category (e.g., Genetics or Microbiology).

Protein sequencing with single amino acid resolution discerns peptides that discriminate tropomyosin proteoforms

Natchanon Sittipongpittaya^{1#}, Kenneth A. Skinner^{4##}, Erin D. Jeffery¹, Emily F. Watts¹, Gloria M. Sheynkman^{1,2,3,*}

¹Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, Virginia 22903, United States

²Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, Virginia 22903, United States

³UVA Comprehensive Cancer Center, University of Virginia, Charlottesville, Virginia 22903, United States

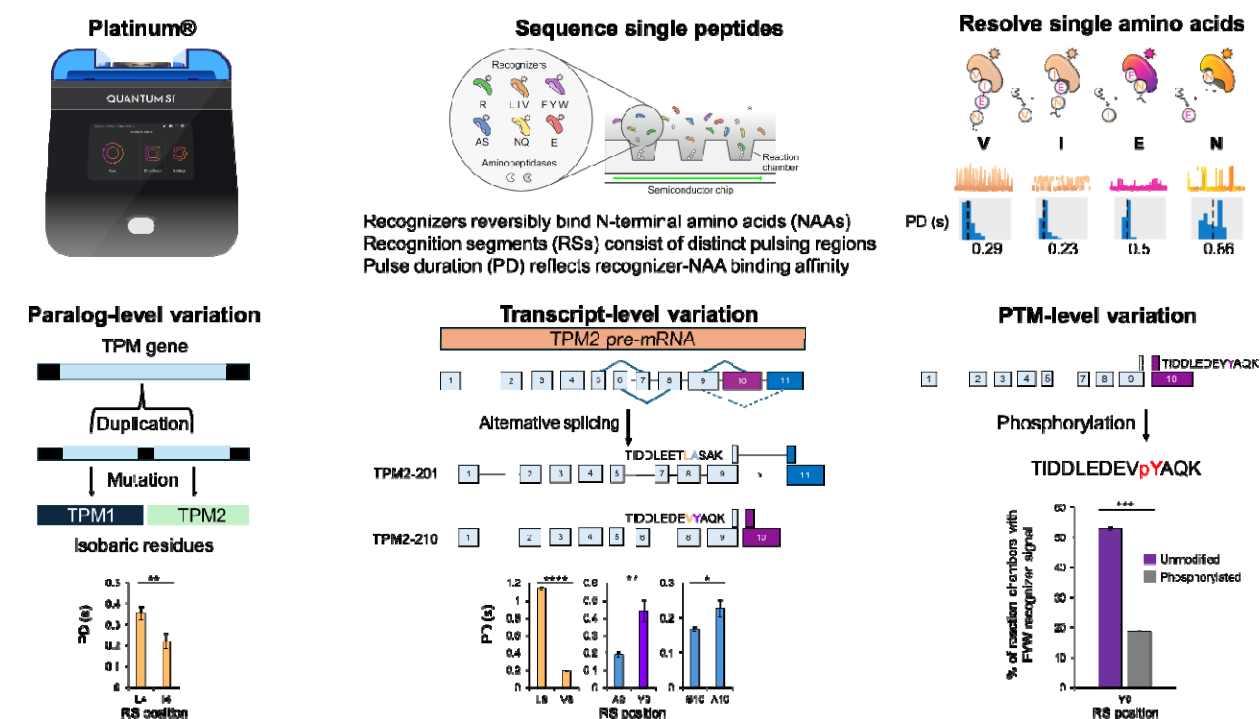
⁴Quantum-Si Incorporated, 29 Business Park Drive, Branford, Connecticut 06405, United States

#Equal contribution

*Co-corresponding author email addresses: kskinner@quantum-si.com; gs9yr@virginia.edu

Abstract

Protein variants of the same gene—proteoforms—can have high molecular similarity yet exhibit different biological functions. Thus, identifying unique peptides that unambiguously map to proteoforms can provide crucial biological insights. In humans, four human tropomyosin (TPM) genes produce similar proteoforms that can be challenging to distinguish with standard proteomics tools. For example, TPM1 and TPM2 share 85% sequence identity, with amino acid substitutions that play unique roles in muscle contraction and myopathies. In this study, we evaluated the ability of the recently released Platinum single-molecule protein sequencer to detect proteoform-informative peptides. Platinum employs fluorophore-labeled recognizers that reversibly bind to cognate N-terminal amino acids (NAAs), enabling polypeptide sequencing within nanoscale apertures of a semiconductor chip that can accommodate single peptide molecules. As a proof of concept, we evaluated the ability of Platinum to distinguish three main types of proteoform variation: paralog-level, transcript-level, and post-translational modification (PTM). We distinguished paralogous TPM1 and TPM2 peptides differing by a single isobaric residue (leucine/isoleucine). We also distinguished tissue-specific TPM2 spliceforms. Notably, we found that a phosphotyrosine-modified peptide displayed reduced recognizer affinity for tyrosine, showing sensitivity to PTMs. This study paves the way for the targeted detection of proteoform biomarkers at the single molecule level.



Key Words

Amino acid variant, single-molecule peptide sequencing, tropomyosin, genetic variants, alternative splicing, phosphorylation, recognizer-based sequencing, proteoforms, isobaric peptides, proteogenomics

Introduction

Proteins serve as the vital molecular bridge between genotype and phenotype, making their functional characterization critical for complete understanding of biological processes¹. Protein function, however, is not fully encoded by the genome. Alternative splicing (AS) of pre-mRNA transcripts can produce distinct protein isoforms^{2,3}. Genes can also be regulated at the post-transcriptional level, and proteins often undergo post-translational modifications (PTMs) that can alter function⁴. The protein products of genetic, post-transcriptional, and post-translational variation—collectively termed proteoforms⁵—can exhibit strikingly diverse functions that give rise to phenotypes in human health and disease^{2,4}. Thus, detecting these proteoforms is key for understanding the molecular link between protein function and phenotype.

Characterizing proteoform expression can present challenges. Currently, the existence of most proteoforms is inferred from mRNA transcripts⁶ and putative PTM sites⁷. However, not all mRNA transcripts are translated into protein isoforms⁸, and some transcripts are degraded without translation^{9–11}. In addition, even if putative PTM sites for all proteoforms are known, the exact proteoforms in a sample cannot be identified unless PTMs are detected and localized to specific sites¹². Indeed, the vast majority of predicted proteoforms have yet to be validated at the peptide or protein level^{13,14}. To bridge this gap, it is necessary to leverage a range of analytical technologies with varying capabilities, both independently and in combination.

Traditional antibody-based assays cannot distinguish proteoforms unless antibodies with high proteoform specificity are available^{15,16}. While mass spectrometry (MS)-based methods have benefits such as high throughput and versatility, they can miss proteoform-informative peptides due data-dependent acquisition (DDA) selection bias towards the most intense peptide ions¹⁷. In addition, resolving peptides with similar physicochemical properties is challenging with size- and charge-based chromatographic methods¹⁸. Recently, there has been considerable progress in identifying peptides that uniquely indicate specific proteoforms^{19,20}, an important advancement for discerning proteoforms with high amino acid sequence identity. However, to fully characterize proteoforms, it will be necessary to employ new and orthogonal detection methods capable of addressing some of the drawbacks of current methods.

Recently, several single molecule protein sequencing approaches have emerged with the potential to detect subtle amino acids and variations^{21–23}. Recognizer-based sequencing, also referred to as next-generation protein sequencing (NGPS), was the first of these to be commercialized with the release of the Platinum instrument (Quantum-Si). Therefore, we sought to evaluate the feasibility of distinguishing proteoform-informative peptides using NGPS on Platinum. For this study, we chose a model proteoform family—tropomyosins (TPMs)—that

shares high amino acid sequence similarity (**Figure 1**) and represents diversity at the levels of genetic polymorphisms, RNA splice variants, and PTMs³⁰.

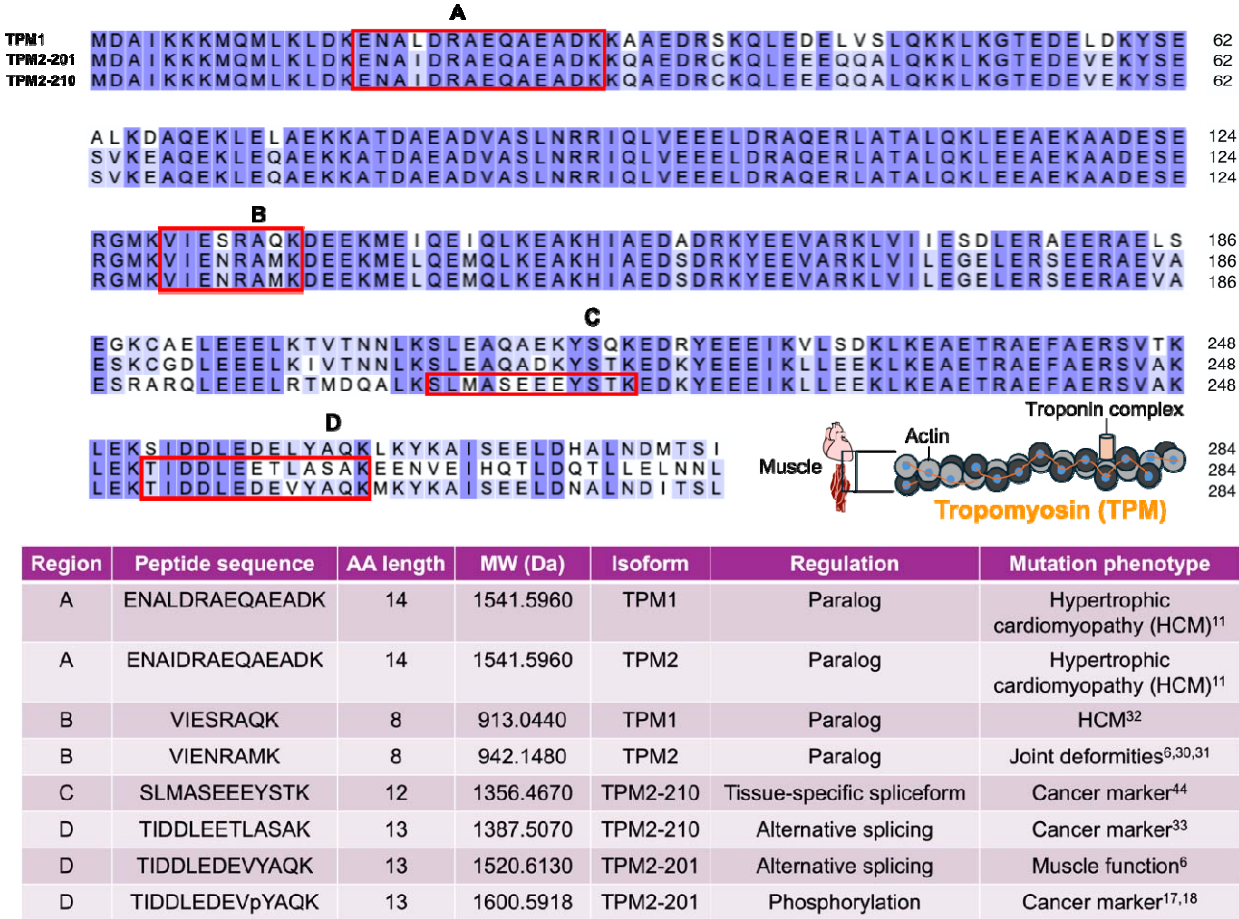


Figure 1 Sequence alignment of TPM1 and TPM2 isoforms and relevance of amino acid mutations to pathology. The primary structures of canonical TPM1 and two TPM2 spliceforms are shown. TPM and troponin associate with actin to form the contractile apparatus. Region A corresponds to a N-terminal region that contains isobaric peptides that differ by leucine (Leu; L) or isoleucine (Ile; I). Mutations in these paralogs have been linked with hypertrophic cardiomyopathy (HCM)²⁴. Region B delineates an internal eight amino acid long region within TPM1/2 paralogs. Mutations within this region are also found in HCM and joint deformities due to Sheldon-Hall syndrome and distal arthrogryposis^{25–27}. Region C highlights an isoform-informative peptide (SLMASEEEYSTK) generated via alternative splicing. Region D maps to an alternatively spliced C-terminal region of TPM2-201 and TPM2-210. While TIDDLEETLASAK (TPM2-201) and TIDDLEDEVYAQK (TPM2-210) share the same amino acid length, TIDDLEDEVYAQK is subject to phosphorylation on a single tyrosine residue^{28,29}.

Filament proteins such as TPMs are among the most highly conserved proteins in evolution, playing critical roles in cellular architecture, cytoskeletal movement, and compartmental trafficking³⁰. TPMs are alpha-helical coiled coil dimers that regulate the stability of actin filaments in muscle and non-muscle cells³¹. Via alternative splicing, tissue-specific promoter usage, and different poly (A) addition sites, the four human TPM genes (**Supplementary Figure 1**) collectively produce more than 40 mRNA variants that can be translated as tissue-specific proteins in skeletal and non-skeletal tissue types^{30,32}.

Missense mutations in TPMs are linked to cardiac diseases such as hypertrophic cardiomyopathy (HCM)²⁴ and skeletal muscle diseases^{24,33–35} (**Figure 1**). These mutations may modulate interactions of TPMs with other filament proteins, impacting the contractile apparatus^{36–38}. Altered TPM expression and phosphorylation have also been linked with dilated cardiomyopathy (DCM) and cancer phenotypes^{28,29}. Because TPMs are subject to extensive variation at the genetic, post-transcriptional, and post-translational levels³⁹, the functional diversity and high sequence similarity of TPM proteoforms presents a daunting challenge with current analytical methods⁴⁰.

Therefore, in this study, we sought to determine whether Platinum could distinguish between diverse TPM proteoforms. We focused on three major types of variants: isobaric peptides from TPM1/2 gene paralogs, peptides derived from tissue-specific TPM2 spliceforms, and phosphotyrosine-modified variants from unmodified peptides. In the process, we leveraged key features of the Platinum sequencing data type: characteristic pulse durations (PDs), which distinguish structurally similar NAAs, and the order of discrete recognition segments (RSs), which pinpoint amino acid location. Together, these measurements produce “kinetic signatures” that provide information on both amino acid identity and sequential order. Given that sequence variations within our model system correspond to regions relevant to TPM pathophysiology (**Figure 1**), we envision broader applications of NGPS to the targeted detection of proteoform-informative peptides from biological samples.

Materials and Methods

In silico digestion and peptide selection criteria

Protein sequences of canonical TPM1 and TPM2 spliceforms⁴¹ (**Figure 1**) were extracted from GENCODE (v43). In-silico LysC⁴² digestion of TPM sequences was performed using a custom R script leveraging the R package cleaver. The resulting peptide sequences were mapped to their corresponding genomic coordinates using PoGo and visualized as a UCSC browser track to identify peptides that are specific to TPM1 (ENALDRAEQAEADK and VIESRAQK) or TPM2 (ENALDRAEQAEADK and VIENRAMK), shared between spliceforms of TPM2 (VIENRAMK), and specific to certain spliceforms of TPM2 (SLMASEEEYSTK, TIDDLEETLASAK, and TIDDLEDEVYAQK).

Peptides that are paralog-specific, shared between TPM2 spliceforms, and TPM2 spliceform-specific were filtered according to the following criteria: the peptide contains a C-terminal lysine (Lys; K), the peptide contains ≥ 3 amino acids that are cognate to the NAA recognizers, the peptide sequence can be recognized by ≥ 3 different recognizers (**Supplementary Figure 3**), and

the peptide contains 5-25 amino acids. These criteria are important for successful sequencing of peptides by NGPS (Library Preparation Kit -Lys-C Data Sheet and Platinum Analysis Software Data Sheet).

Peptide synthesis

Eight peptides identified from the *in silico* peptide screening (**Figure 1**) were synthesized by JPT Peptide Technologies, Inc. (Berlin, Germany) with C-terminal carboxylic acid and azido-lysine modifications (1 mg, >90% purity, lyophilized). Peptides were reconstituted in 50% acetonitrile to a concentration of 5 mM each and stored at -80°C.

Synthetic phosphopeptides and unphosphorylated peptides with C-terminal azido-lysine modifications were synthesized by Genscript (Piscataway, NJ, USA) (4 mg, >90% purity, lyophilized). The peptides were reconstituted in dimethylformamide to a concentration of 5 mM each and stored at -80°C.

LC-MS/MS analysis

Peptide separation was performed using nanoflow high-performance liquid chromatography (HPLC) on a Dionex Ultimate 3000 system (Thermo Fisher Scientific, Bremen, Germany). Peptides were initially loaded onto an Acclaim PepMap 100 trap column (300 µm × 5 mm, 5 µm C18), followed by gradient elution through an Acclaim PepMap 100 analytical column (75 µm × 25 cm, 3 µm C18) for enhanced separation. Mass spectrometry (MS) analysis was conducted using an Orbitrap Eclipse Tribrid mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with the Orbitrap Eclipse Tune (version 4.0.4091) and Xcalibur software (version 4.5.445.18) for data acquisition and analysis.

Preparation of linker-conjugated peptide libraries

Synthetic peptides with C-terminal azido-lysine modifications were conjugated to linker molecules via strain-induced click conjugation⁴³. Peptides were diluted to a final concentration of 50 µM in 100 mM HEPES, pH 8.0 (20% acetonitrile) and incubated overnight with 2 µM linker at 37 °C. The resulting conjugated peptide libraries were stored at -20 °C until sequencing.

Peptide sequencing on Platinum

Experiments were conducted in accordance with Library Preparation Kit -Lys-C Data Sheet and Platinum Instrument and Sequencing Kit V2 Data Sheet (February 27, 2024).

Library Preparation Kit -Lys-C Data Sheet link: <https://www.quantum-si.com/resources/product-data-sheets/library-preparation-kit-lys-c-data-sheet/>

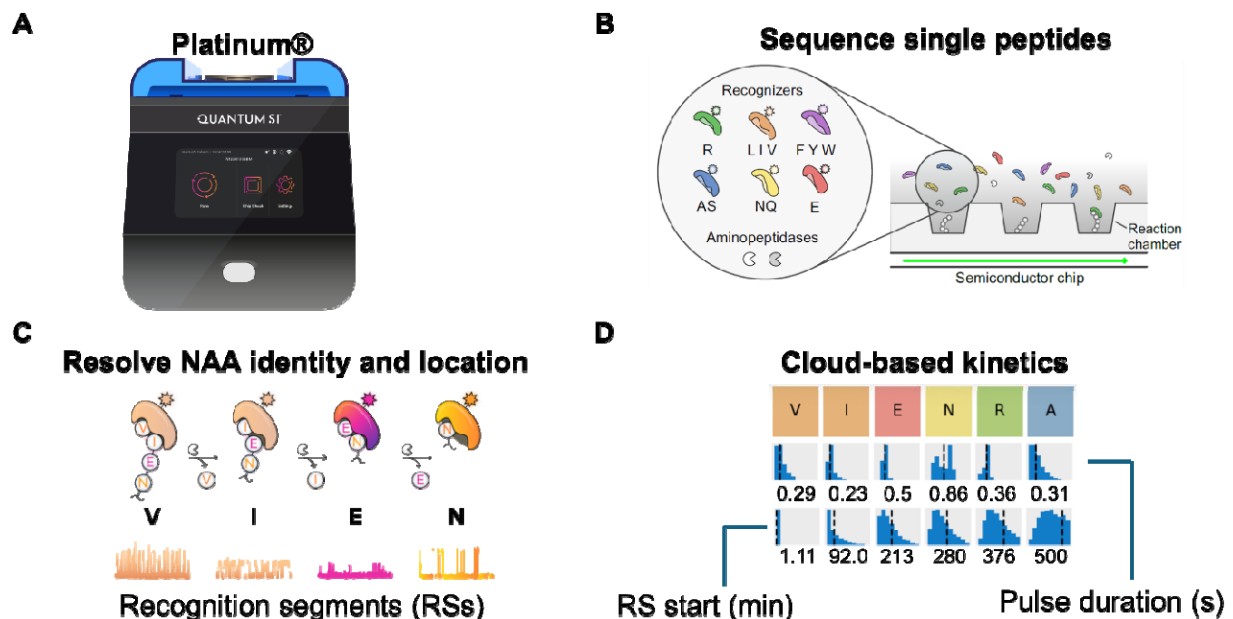
Platinum Instrument and Sequencing Kit V2 Data Sheet link: <https://www.quantum-si.com/resources/product-data-sheets/platinum-instrument-and-sequencing-kit-v3-data-sheet/>

Briefly, conjugated peptides are immobilized in nanoscale reaction chambers on a semiconductor chip (**Figure 2A**) for exposure to a mixture of freely diffusing NAA recognizers and

aminopeptidases. The mixture consisted of NAA recognizers that target 12 of 20 canonical AAs (**Figure 2B**). During on-chip sequencing, NAA recognizers reversibly bind cognate NAAs, producing recognition segments (RSs) that are captured by the semiconductor chip (**Figure 2C**). To carry out the sequencing process, aminopeptidases cleave the peptide bond and expose the subsequent NAA for recognition. After 10 hours of runtime, sequencing data is transferred to the Platinum Analysis Software.

Cloud-based analysis of sequencing data; analysis versions

Primary Analysis v2.5.1 and Peptide Alignment v2.5.2 were used for cloud-based analysis of sequencing data. Details can be found in the Platinum Analysis Software Data Sheet (February 2, 2024). Upon completion of sequencing run, the fluorescent signal, derived from the repeated on-off binding events of the recognizers and NAAs, is securely transferred to the Platinum Analysis Software. The Primary Analysis workflow is the first step in processing data which characterizes the apertures across the chip based on peptide loading, recognizer activity, recognizer reads, and recognizer read lengths (**Supplementary Figure 3**). For the Peptide Alignment workflow, reads are aligned to the reference profile based on the correspondence of observed recognition segments to the expected reference profile, using recognizer identity.



Results and Discussion

We first sought to identify proteoform-specific peptides in TPM1 and 2 to serve as the focus of this study. We performed an *in silico* digestion of TPM1 and TPM2 with Lys-C, an enzyme that cleaves peptide bonds at the C-terminal side of lysine (K) residues⁴². We then selected eight peptides that correspond to distinct regions along TPM1/2 primary structure (**Figure 1**). In addition, amino acid mutations in these peptides have been found to be relevant to TPM pathophysiology^{24–29} (**Figure 1 and Supplementary Figure 2**).

To discriminate gene paralogs at the peptide level, we tested two sets of paralogous, isobaric peptides on Platinum: ENALDRAEQAEADK (TPM1)/ENAI DRAEQAEADK (TPM2) (**Figure 1 Peptide Pair A, Figure 3A**) and VIESRAQK (TPM1)/VIENRAMK (TPM2) (**Figure 1 Peptide Pair B, Figure 3C**). ENALDRAEQAEADK and ENAI DRAEQAEADK are identical in amino acid sequence except for the 4th position from the N-terminus, in which there is an isobaric Leu/Ile substitution (**Figure 1 Peptide Pair A, Figure 3A**). In other words, they share the same length and molecular weight, and so specifically detecting each peptide requires differentiation of signals from Leu and Ile. We conjugated the peptides and mixed the resulting linker-derivatized peptides in an equimolar ratio for loading on both flow cells of the semiconductor chip (**Methods**). For comparison of ENALDRAEQAEADK and ENAI DRAEQAEADK, the first three amino acids (Glu1/E1, Asn2/N2, Ala3/A3) are identical and each can be measured via distinct recognizers (**Figure 2B**). The recognizer for Leu, Ile, and Val has the highest affinity for N-terminal Leu and this preference is expected to result in markedly longer PD for LDR compared to IDR in these two peptides²¹.

Indeed, we observed a higher average PD for Leu (0.35 s) compared to Ile (0.23 s) (**Figure 3B**). These results indicate a single recognizer distinguishes isobaric residues that differ at position 4 from the N-terminus, thus confidently discriminating these paralog-specific peptides.

Interestingly, we observed similar trends of highly specific peptide recognition with the paralogous peptides VIENRAMK and VIESRAQK, in which all residues are the same except for the Asn to Ser substitution in position 4 and the Met to Gln substitution in position 7 (**Figure 1 Peptide Pair B, Figure 3C**). While all matched residues between the two peptides exhibit similar PD profiles, the Asn to Ser substitution at position 4 was distinguished via the dye call for each recognizer (**Figure 3C**). In addition, the recognizers for N and S elicited different PDs for their cognate residues, providing an additional metric for distinguishing each residue at this position. Indeed, a ~5-fold longer PD was detected for Asn relative to Ser (**Figure 3D**).

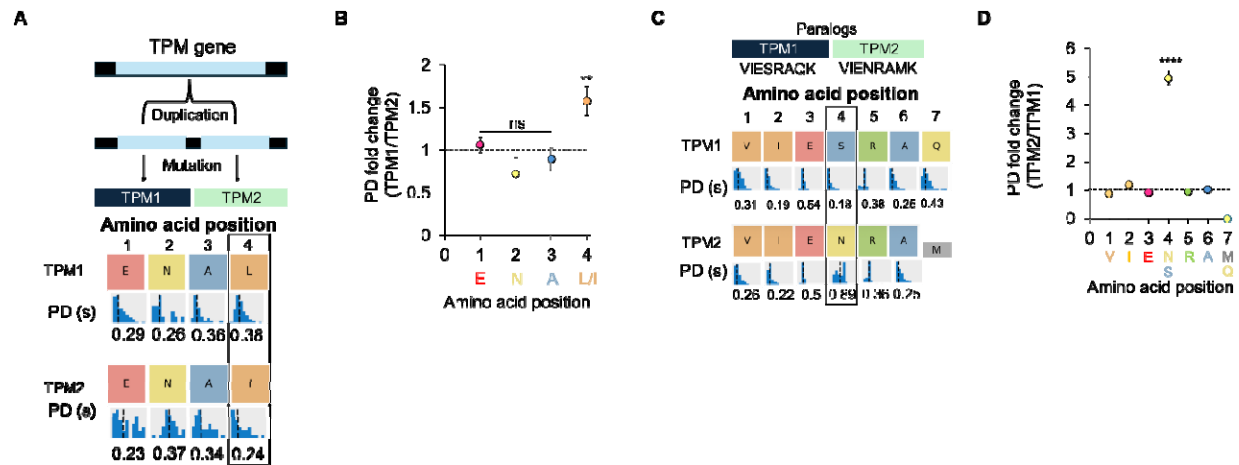


Figure 3. Kinetic signatures discern the identity and amino acid ordering of paralog-derived peptides. (A) Gene duplication events produce TPM1/TPM2 paralogs, wherein a single isobaric L to I substitution differentiates ENALDRAEQAEADK from ENAIDRAEQAEADK (top). An example kinetic signature snapshot from cloud-based analysis is shown below the amino acid positions (bottom). Colors of the boxes indicate the recognizer identity (see Figure 1B). Kinetic signatures such as PD represent the statistical distribution of kinetic data for all pulses associated with a specific residue. (B) Plot of the fold change in the average pulse duration (PD) between NAA measurements of ENAL versus ENAI, for the first 4 residues. Error bars indicate standard deviation from three experiments conducted on independent flow cells. Asterisks indicate p-values relative to a PD fold change of 1. (C) Gene duplication events produce TPM1/TPM2 paralogs that can be differentiated at position 4 of TPM1 peptide VIESRAQK and TPM2 peptide VIENRAMK. (D) Plot of the fold change PDs for VIENSRAQ versus VIENRAM. A ~5-fold higher PD was detected for the N in position 4 of VIENRAMK, relative to S in VIESRAQK, which was statistically significant, when compared to the PD fold change to all other positions. Data (*mean \pm S.D.) are representative of three independent experiments, conducted on independent flow cells (ns $p > 0.05$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$; two-tailed t-test). PD=Pulse Duration

In addition to paralogs, the TPM gene family also produces alternatively spliced isoforms that are tissue specific and functionally distinct. In TPM2, exon 6 or 7 is included, but not both, with the same mutually exclusive splicing of exons found for exon 10 and 11. We found that SLMASEEEYSTK maps to a specific splice junction that overlaps exon 6, and thus could inform the presence of the spliceform TPM2-210, which has been found to be expressed in non-skeletal muscle⁴⁴⁻⁴⁷ (Figure 1 Peptide Pair C, Figure 4A). From the sequencing of SLMASEEEYSTK, PDs were generated for all amino acids within the peptide, except for the Met at position 3 (Figure 4B).

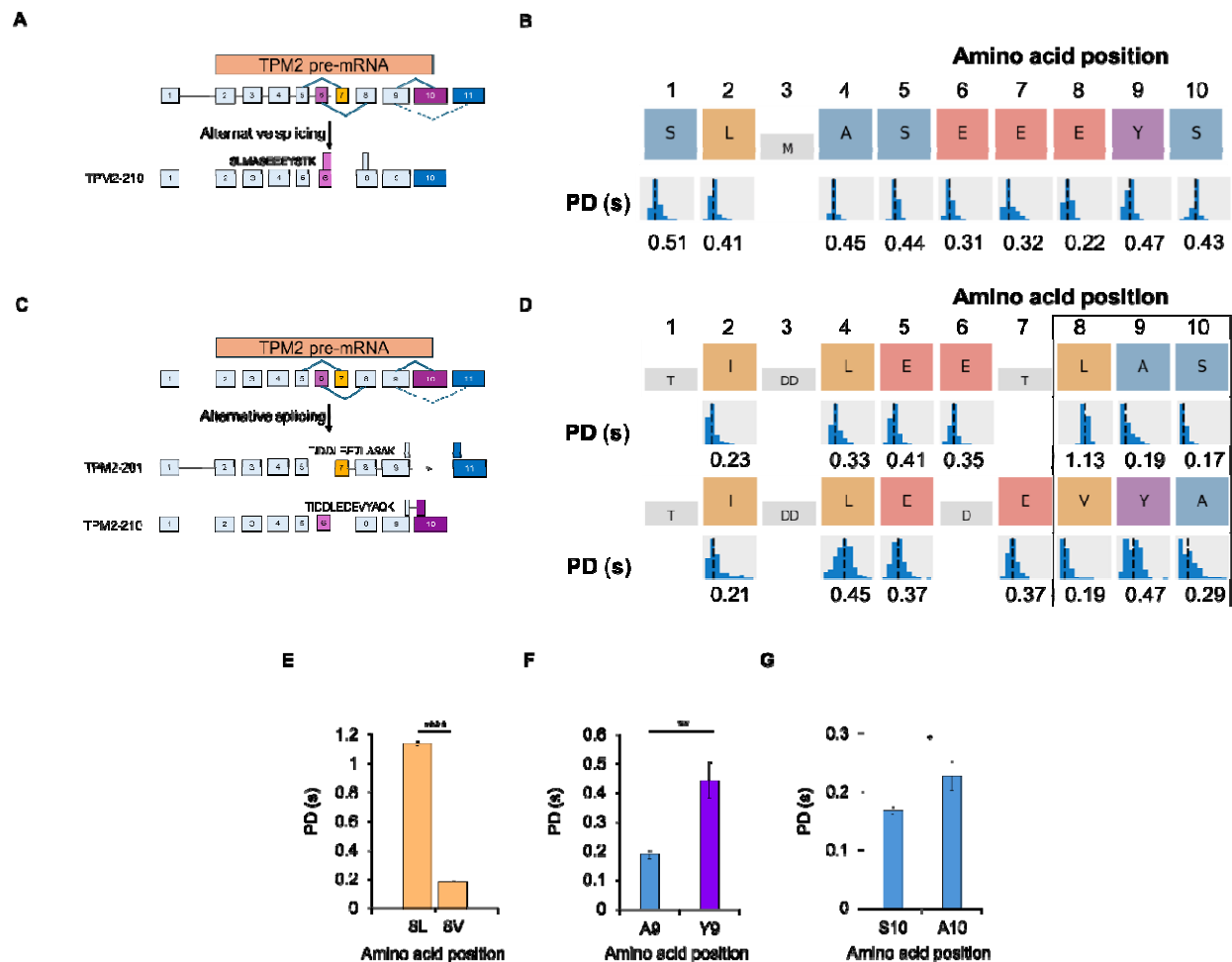


Figure 4. Kinetic signatures discern the identity and temporal order of TPM2 peptides derived from alternative splicing. (A) Alternative splicing of TPM2 mRNA generates a tissue-specific spliceform (containing Exon 6) to which SLMASEEEYSTK specifically maps. (B) Sequence alignments for SLMASEEEYSTK. An example kinetic signature snapshot from cloud-based analysis is shown below the amino acid positions. Colors of the boxes indicate the recognizer identity (see **Figure 1B**). Kinetic signatures such as PD represent the statistical distribution of kinetic data for all pulses associated with a specific residue. (C) Alternative splicing of TPM2 mRNA differentiates exon-specific peptides TIDDLLEETLASAK (Exon 10) and TIDDLLEDEVYAQK (Exon 11). (D) Sequence alignments for peptides TIDDLLEETLASAK and TIDDLLEDEVYAQK. (E) LIV recognizer elicits differential PD profiles for similar aliphatic residues. A ~9-fold difference in PD was observed for L/V. (F) Distinct recognizers for A and Y discern position 9. (G) The A/S recognizer enables discernment via differential PDs at position 10. Data (*mean \pm S.D.) are representative of three independent experiments, conducted on independent flow cells (ns $p > 0.05$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$; two-tailed t-test). PD=Pulse Duration

We also found a peptide pair that discriminates exon 10 from exon 11, with exon-specific peptides of the same length: TIDDLEETLASAK (exon 11) and TIDDLEDEVYAQK (exon 10) (**Figures 1 Peptide Pair D, 4C**). The combination of residue position as well as recognizer identity and PD led to a highly specific discrimination of the peptides (**Figure 4D**). In particular, residues in positions 8-10 were clearly distinguished by recognizer and PD patterns. For position 8—in an additional demonstration of the ability of the LIV recognizer to discern aliphatic amino acids (**Figure 3B**)—we observed a ~9-fold higher PD for the Leu relative to valine (Val; V) (**Figures 4E**). In position 9, Tyr and Ala residues were clearly distinguished with distinct recognizers (**Figures 4F**). In further support of the ability of a single recognizer to elicit differential PD profiles, we observed a higher PD for Ala relative to Ser in position 10 (**Figures 4G**). Taken together, these results demonstrate specific sequencing of splice junction-specific peptides with Platinum, and that Platinum differentiates peptides based on both structurally similar and distinct amino acid side chains.

Finally, we tested the ability of Platinum to distinguish peptidoforms⁴⁸ arising from PTMs—a major source of proteoform variation. One of the benefits of single-molecule sequencing is discrete binding events between each recognizer and its cognate amino acid, which is a highly sensitive modality of detection. We found that the previously sequenced peptide TIDDLEDEVYAQK (**Figure 4**), which is specific to exon 10 of TPM2, contains an annotated phosphorylation at the Tyr in position 10⁴⁹. The addition of a phosphate group significantly changes the charge and topology of Tyr, weakening binding by the recognizer for bulky aromatic residues (**Figures 1B**).

We separately conjugated TIDDLEDEVYAQK and TIDDLEDEVpYAQK (**Figure 5A**), then loaded each peptidoform for on-chip sequencing in independent flow cells. As expected, we observed a significant decrease in percent of total nanoapertures containing Tyr RSs for the phosphorylated form relative to the unmodified form, while the percent of nanoapertures containing LIV and E RSs were similar between the two groups (**Figure 5B**). These results indicate that Platinum sequencing is sensitive to the presence of PTMs; thus, kinetic signatures can be used to infer the presence of PTMs at single amino acid resolution.

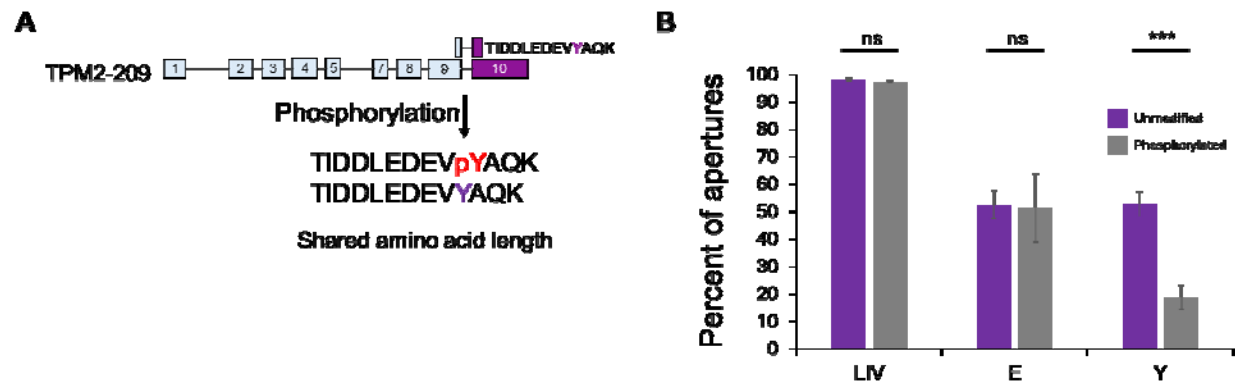


Figure 5. YFW recognizer generates Y-specific recognition events, enabling differentiation from phosphotyrosine-modified peptidofoms. (A) Schema showing the exon-specific peptide TIDDLEDEVYAQK, which contains a known phosphosite at the tyrosine residue. (B) Bar plot showing the difference in percent of apertures between the recognizers (LIV, E, and FYW) in the unmodified versus phosphorylated version (TIDDLEDEVpYAQK). Data (mean \pm S.D.) are representative of three independent experiments, conducted on independent flow cells (ns $p > 0.05$, *** $p < 0.001$; Welch two-sample t-test).

Conclusions

Analyzing the proteome is essential for a comprehensive understanding of biological processes. However, several fundamental challenges hinder the ability to fully characterize proteomic diversity. These challenges stem from the complexity of the proteome, which could contain over 1 million distinct proteoforms^{1,14}.

Existing methods for detecting proteoform-informative peptides, such as mass spectrometry, can struggle with distinguishing closely related variants, particularly those that differ by single amino acid substitutions or PTMs. These methods can have other limitations, including high costs and detection thresholds, as well as the requirement for specific expertise to operate and interpret MS experiments. In contrast, by directly sequencing individual peptides in an accessible benchtop format, Platinum has the potential to overcome some of the limitations of traditional methods, serving as an orthogonal tool for proteomic discovery.

In this study, we sought to determine the ability of the benchtop Platinum instrument to distinguish proteoform-informative peptides derived from the human TPM gene family. Using this single-molecule protein sequencing technology, we identify key variations at the paralog, transcript, and PTM levels. While prior work²¹ laid the foundation for the detection of amino acid variation using Platinum, this study represents the first application of Platinum for sequencing proteoform-informative peptides.

We demonstrate the ability of Platinum to differentiate between paralogous TPM1 and TPM2 peptides that differ by a single amino acid substitution, as well as to sequence tissue-specific TPM2 splice variants. As TPMs are subject to post-transcriptional regulation (e.g. long non-coding RNA and antisense transcripts)^{11,50}, there is a need to validate TPM expression at the

peptide/protein level. In future studies with biological samples, a pan-TPM antibody that recognizes a conserved site on each TPM isoform could be used for upfront enrichment, followed by Platinum sequencing to identify peptides that uniquely map to each TPM variant. More generally, given its orthogonal chemistry, Platinum can provide additional evidence to support peptide identifications via DDA from MS (**Supplementary Figure 4A and B**).

Our findings also illustrate the sensitivity of Platinum to phosphotyrosine modifications at the single amino acid level. Given this capability, future studies could use Platinum to perform target identification for phospho-specific antibodies⁵¹. Detecting phosphorylation is particularly challenging due to the transient nature of these modifications and the complex interplay of signaling networks, making high-resolution methods essential for capturing these dynamic processes⁵².

There are several limitations to this study. First, we chose to focus on proteoform-informative synthetic peptides as a model system for this foundational proof-of-concept evaluation. Thus, future studies will be required to extend these findings to biological samples. Second, because we used either 1:1 mixtures of peptides or individual peptides loaded in independent flow cells (Y and pY), future studies with complex mixtures will be required to simulate conditions typically obtained from proteomic studies. Beyond this study, several innovations are needed to expand the biological applications of Platinum, including the development of additional recognizers to expand sequencing coverage. Recently, Quantum-Si released a combined Asp/Glu recognizer; as tropomyosin proteoforms contain an abundance of acidic residues (**Figure 1 and Supplementary Figure 5**), additional peptides could likely be resolved by differentiating acidic sites with the latest sequencing kit.

Broadly, the application of Platinum for proteoform differentiation is generalizable to other proteoform families. While TPMs have been greater than 500 million years in the making³⁶, new TPM discoveries are on the horizon, with not only mammals but also six TPM genes that have been found in zebrafish, a popular animal model for gene expression studies⁵³. In addition, TPMs are the primary allergens in shellfish, causing food allergies that affect 2% of the US adult population^{54,55}. Hence, detecting TPM peptides that elicit food allergies could be of great interest to the food science community⁴⁰.

More generally, NGPS on Platinum could integrate with existing proteogenomic approaches to identify immunopeptides derived from noncoding regions^{56,57}, enabling detection of functional biomolecules often overlooked by conventional methods. Further, combining NGPS with genomic and transcriptomic data could provide a more comprehensive view of proteoform diversity, including low-abundance variants and post-translational modifications. This multi-omics strategy will lead to an improved understanding of the proteome in health and disease.

Author contributions

Conception or project design: NS, KAS, EDJ, GMS

Data acquisition: NS, KAS, EDJ, GMS

Data analysis and interpretation of results: NS, KAS, EDJ, EFW, GMS

Original draft preparation: NS, KAS, EDJ, GMS

All authors reviewed the results and approved the final version of the manuscript.

Disclosures

KAS is an employee and shareholder of Quantum-Si. No other authors have any disclosures.

Acknowledgments

We thank members of the Sheynkman Lab, particularly Leon Sheynkman (University of Virginia) for technical assistance. Also, we thank Cheng Man Lun, Meredith Carpenter, and Brian Reed for helpful discussions (Quantum-Si).

References

- (1) Aebersold, R.; Agar, J. N.; Amster, I. J.; Baker, M. S.; Bertozzi, C. R.; Boja, E. S.; Costello, C. E.; Cravatt, B. F.; Fenselau, C.; Garcia, B. A.; Ge, Y.; Gunawardena, J.; Hendrickson, R. C.; Hergenrother, P. J.; Huber, C. G.; Ivanov, A. R.; Jensen, O. N.; Jewett, M. C.; Kelleher, N. L.; Kiessling, L. L.; Krogan, N. J.; Larsen, M. R.; Loo, J. A.; Ogorzalek Loo, R. R.; Lundberg, E.; MacCoss, M. J.; Mallick, P.; Mootha, V. K.; Mrksich, M.; Muir, T. W.; Patrie, S. M.; Pesavento, J. J.; Pitteri, S. J.; Rodriguez, H.; Saghatelian, A.; Sandoval, W.; Schlüter, H.; Sechi, S.; Slavoff, S. A.; Smith, L. M.; Snyder, M. P.; Thomas, P. M.; Uhlén, M.; Van Eyk, J. E.; Vidal, M.; Walt, D. R.; White, F. M.; Williams, E. R.; Wohlschlager, T.; Wysocki, V. H.; Yates, N. A.; Young, N. L.; Zhang, B. How Many Human Proteoforms Are There? *Nat. Chem. Biol.* **2018**, *14* (3), 206–214. DOI: 10.1038/nchembio.2576.
- (2) Kjer-Hansen, P.; Weatheritt, R. J. The Function of Alternative Splicing in the Proteome: Rewiring Protein Interactomes to Put Old Functions into New Contexts. *Nat. Struct. Mol. Biol.* **2023**, *30* (12), 1844–1856. DOI: 10.1038/s41594-023-01155-9.
- (3) Kelemen, O.; Convertini, P.; Zhang, Z.; Wen, Y.; Shen, M.; Falaleeva, M.; Stamm, S. Function of Alternative Splicing. *Gene* **2013**, *514* (1), 1–30. DOI: 10.1016/j.gene.2012.07.083.
- (4) Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J. Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications. *Angew. Chem. Int. Ed* **2005**, *44* (45), 7342–7372. DOI: 10.1002/anie.200501023.
- (5) Smith, L. M.; Kelleher, N. L.; Consortium for Top Down Proteomics. Proteoform: A Single Term Describing Protein Complexity. *Nat. Methods* **2013**, *10* (3), 186–187. DOI: 10.1038/nmeth.2369.
- (6) Reixachs-Solé, M.; Eyras, E. Uncovering the Impacts of Alternative Splicing on the Proteome with Current Omics Techniques. *Wiley Interdiscip. Rev. RNA* **2022**, *13* (4), e1707. DOI: 10.1002/wrna.1707.
- (7) Ramazi, S.; Zahiri, J. Posttranslational Modifications in Proteins: Resources, Tools and Prediction Methods. *Database (Oxford)* **2021**, *2021*. DOI: 10.1093/database/baab012.
- (8) Tress, M. L.; Abascal, F.; Valencia, A. Alternative Splicing May Not Be the Key to Proteome Complexity. *Trends Biochem. Sci.* **2017**, *42* (2), 98–110. DOI: 10.1016/j.tibs.2016.08.008.
- (9) Weatheritt, R. J.; Sterne-Weiler, T.; Blencowe, B. J. The Ribosome-Engaged Landscape of Alternative Splicing. *Nat. Struct. Mol. Biol.* **2016**, *23* (12), 1117–1123. DOI: 10.1038/nsmb.3317.
- (10) Lewis, B. P.; Green, R. E.; Brenner, S. E. Evidence for the Widespread Coupling of Alternative Splicing and Nonsense-Mediated mRNA Decay in Humans. *Proc Natl Acad Sci USA* **2003**, *100* (1), 189–192. DOI: 10.1073/pnas.0136770100.
- (11) Huang, G.-W.; Zhang, Y.-L.; Liao, L.-D.; Li, E.-M.; Xu, L.-Y. Natural Antisense Transcript TPM1-AS Regulates the Alternative Splicing of Tropomyosin I through an Interaction with RNA-Binding Motif Protein 4. *Int. J. Biochem. Cell Biol.* **2017**, *90*, 59–67. DOI: 10.1016/j.biocel.2017.07.017.

- (12) Potel, C. M.; Lemeer, S.; Heck, A. J. R. Phosphopeptide Fragmentation and Site Localization by Mass Spectrometry: An Update. *Anal. Chem.* **2019**, *91* (1), 126–141. DOI: 10.1021/acs.analchem.8b04746.
- (13) Ruggles, K. V.; Tang, Z.; Wang, X.; Grover, H.; Askenazi, M.; Teubl, J.; Cao, S.; McLellan, M. D.; Clauser, K. R.; Tabb, D. L.; Mertins, P.; Slebos, R.; Erdmann-Gilmore, P.; Li, S.; Gunawardena, H. P.; Xie, L.; Liu, T.; Zhou, J.-Y.; Sun, S.; Hoadley, K. A.; Perou, C. M.; Chen, X.; Davies, S. R.; Maher, C. A.; Kinsinger, C. R.; Rodland, K. D.; Zhang, H.; Zhang, Z.; Ding, L.; Townsend, R. R.; Rodriguez, H.; Chan, D.; Smith, R. D.; Liebler, D. C.; Carr, S. A.; Payne, S.; Ellis, M. J.; Fenyő, D. An Analysis of the Sensitivity of Proteogenomic Mapping of Somatic Mutations and Novel Splicing Events in Cancer. *Mol. Cell. Proteomics* **2016**, *15* (3), 1060–1071. DOI: 10.1074/mcp.M115.056226.
- (14) Ponomarenko, E. A.; Poverennaya, E. V.; Ilgisonis, E. V.; Pyatnitskiy, M. A.; Kopylov, A. T.; Zgoda, V. G.; Lisitsa, A. V.; Archakov, A. I. The Size of the Human Proteome: The Width and Depth. *Int. J. Anal. Chem.* **2016**, *2016*, 7436849. DOI: 10.1155/2016/7436849.
- (15) Suhre, K.; McCarthy, M. I.; Schwenk, J. M. Genetics Meets Proteomics: Perspectives for Large Population-Based Studies. *Nat. Rev. Genet.* **2021**, *22* (1), 19–37. DOI: 10.1038/s41576-020-0268-2.
- (16) Pietzner, M.; Wheeler, E.; Carrasco-Zanini, J.; Kerrison, N. D.; Oerton, E.; Koprulu, M.; Luan, J.; Hingorani, A. D.; Williams, S. A.; Wareham, N. J.; Langenberg, C. Synergistic Insights into Human Health from Aptamer- and Antibody-Based Proteomic Profiling. *Nat. Commun.* **2021**, *12* (1), 6822. DOI: 10.1038/s41467-021-27164-0.
- (17) Sinitcyn, P.; Richards, A. L.; Weatheritt, R. J.; Brademan, D. R.; Marx, H.; Shishkova, E.; Meyer, J. G.; Hebert, A. S.; Westphall, M. S.; Blencowe, B. J.; Cox, J.; Coon, J. J. Global Detection of Human Variants and Isoforms by Deep Proteome Sequencing. *Nat. Biotechnol.* **2023**, *41* (12), 1776–1786. DOI: 10.1038/s41587-023-01714-x.
- (18) Soga, T. Advances in Capillary Electrophoresis Mass Spectrometry for Metabolomics. *TrAC Trends in Analytical Chemistry* **2023**, *158*, 116883. DOI: 10.1016/j.trac.2022.116883.
- (19) Keerthikumar, S.; Mathivanan, S. Proteotypic Peptides and Their Applications. *Methods Mol. Biol.* **2017**, *1549*, 101–107. DOI: 10.1007/978-1-4939-6740-7_8.
- (20) Korchak, J. A.; Jeffery, E. D.; Bandyopadhyay, S.; Jordan, B. T.; Lehe, M. D.; Watts, E. F.; Fenix, A.; Wilhelm, M.; Sheynkman, G. M. IS-PRM-Based Peptide Targeting Informed by Long-Read Sequencing for Alternative Proteome Detection. *J. Am. Soc. Mass Spectrom.* **2024**. DOI: 10.1021/jasms.4c00119.
- (21) Reed, B. D.; Meyer, M. J.; Abramzon, V.; Ad, O.; Ad, O.; Adcock, P.; Ahmad, F. R.; Alppay, G.; Ball, J. A.; Beach, J.; Belhachemi, D.; Bellofiore, A.; Bellos, M.; Beltrán, J. F.; Betts, A.; Bhuiya, M. W.; Blacklock, K.; Boer, R.; Boisvert, D.; Brault, N. D.; Buxbaum, A.; Caprio, S.; Choi, C.; Christian, T. D.; Clancy, R.; Clark, J.; Connolly, T.; Croce, K. F.; Cullen, R.; Davey, M.; Davidson, J.; Elshenawy, M. M.; Ferrigno, M.; Frier, D.; Gudipati, S.; Hamill, S.; He, Z.; Hosali, S.; Huang, H.; Huang, L.; Kabiri, A.; Kriger, G.; Lathrop, B.; Li, A.; Lim, P.; Liu, S.; Luo, F.; Lv, C.; Ma, X.; McCormack, E.; Millham, M.; Nani, R.; Pandey, M.; Parillo, J.; Patel, G.; Pike, D. H.; Preston, K.; Pichard-Kostuch, A.; Rearick, K.; Rearick, T.; Ribezzi-Crivellari, M.; Schmid, G.;

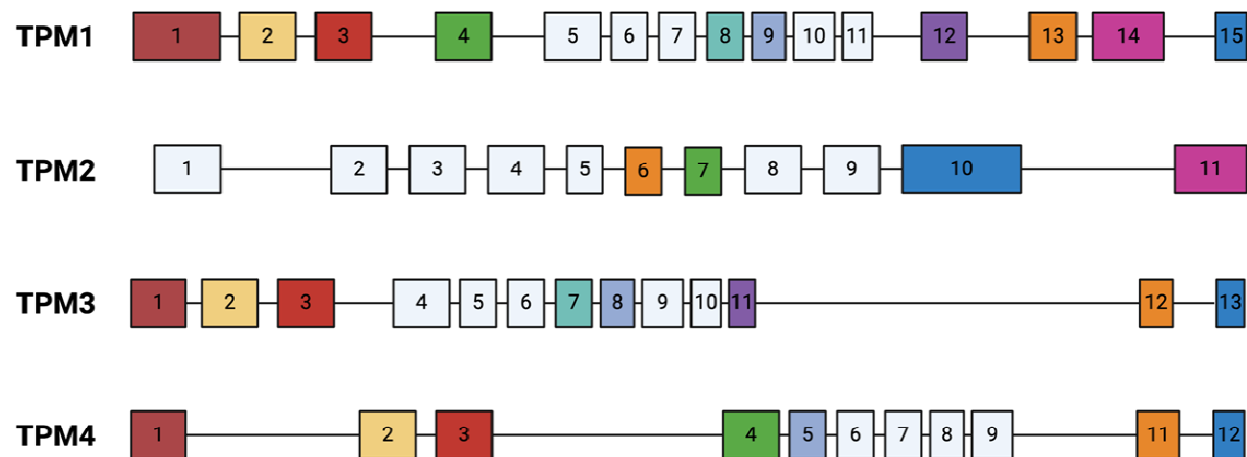
- Schultz, J.; Shi, X.; Singh, B.; Srivastava, N.; Stewman, S. F.; Thurston, T. R.; Thurston, T. R.; Trioli, P.; Tullman, J.; Wang, X.; Wang, Y.-C.; Webster, E. A. G.; Zhang, Z.; Zuniga, J.; Patel, S. S.; Griffiths, A. D.; van Oijen, A. M.; McKenna, M.; Dyer, M. D.; Rothberg, J. M. Real-Time Dynamic Single-Molecule Protein Sequencing on an Integrated Semiconductor Device. *Science* **2022**, 378 (6616), 186–192. DOI: 10.1126/science.abo7651.
- (22) Motone, K.; Kontogiorgos-Heintz, D.; Wee, J.; Kurihara, K.; Yang, S.; Roote, G.; Fox, O. E.; Fang, Y.; Queen, M.; Tolhurst, M.; Cardozo, N.; Jain, M.; Nivala, J. Multi-Pass, Single-Molecule Nanopore Reading of Long Protein Strands. *Nature* **2024**, 633 (8030), 662–669. DOI: 10.1038/s41586-024-07935-7.
- (23) Swaminathan, J.; Boulgakov, A. A.; Hernandez, E. T.; Bardo, A. M.; Bachman, J. L.; Marotta, J.; Johnson, A. M.; Anslyn, E. V.; Marcotte, E. M. Highly Parallel Single-Molecule Identification of Proteins in Zeptomole-Scale Mixtures. *Nat. Biotechnol.* **2018**. DOI: 10.1038/nbt.4278.
- (24) Ly, T.; Pappas, C. T.; Johnson, D.; Schlecht, W.; Colpan, M.; Galkin, V. E.; Gregorio, C. C.; Dong, W.-J.; Kostyukova, A. S. Effects of Cardiomyopathy-Linked Mutations K15N and R21H in Tropomyosin on Thin-Filament Regulation and Pointed-End Dynamics. *Mol. Biol. Cell* **2019**, 30 (2), 268–281. DOI: 10.1091/mbc.E18-06-0406.
- (25) Marttila, M.; Lehtokari, V.-L.; Marston, S.; Nyman, T. A.; Barnerias, C.; Beggs, A. H.; Bertini, E.; Ceyhan-Birsoy, O.; Cintas, P.; Gerard, M.; Gilbert-Dussardier, B.; Hogue, J. S.; Longman, C.; Eymard, B.; Frydman, M.; Kang, P. B.; Klinge, L.; Kolski, H.; Lochmüller, H.; Magy, L.; Manel, V.; Mayer, M.; Mercuri, E.; North, K. N.; Peudenier-Robert, S.; Pihko, H.; Probst, F. J.; Reisin, R.; Stewart, W.; Taratuto, A. L.; de Visser, M.; Wilichowski, E.; Winer, J.; Nowak, K.; Laing, N. G.; Winder, T. L.; Monnier, N.; Clarke, N. F.; Pelin, K.; Grönholm, M.; Wallgren-Pettersson, C. Mutation Update and Genotype-Phenotype Correlations of Novel and Previously Described Mutations in TPM2 and TPM3 Causing Congenital Myopathies. *Hum. Mutat.* **2014**, 35 (7), 779–790. DOI: 10.1002/humu.22554.
- (26) Tajsharghi, H.; Kimber, E.; Holmgren, D.; Tulinius, M.; Oldfors, A. Distal Arthrogryposis and Muscle Weakness Associated with a Beta-Tropomyosin Mutation. *Neurology* **2007**, 68 (10), 772–775. DOI: 10.1212/01.wnl.0000256339.40667.fb.
- (27) Ko, J. M.; Choi, I.-H.; Baek, G.-H.; Kim, K.-W. First Korean Family with a Mutation in TPM2 Associated with Sheldon-Hall Syndrome. *J. Korean Med. Sci.* **2013**, 28 (5), 780–783. DOI: 10.3346/jkms.2013.28.5.780.
- (28) Lundby, A.; Secher, A.; Lage, K.; Nordsborg, N. B.; Dmytriiev, A.; Lundby, C.; Olsen, J. V. Quantitative Maps of Protein Phosphorylation Sites across 14 Different Rat Organs and Tissues. *Nat. Commun.* **2012**, 3, 876. DOI: 10.1038/ncomms1871.
- (29) Huang, C.; Chen, L.; Savage, S. R.; Eguez, R. V.; Dou, Y.; Li, Y.; da Veiga Leprevost, F.; Jaehnig, E. J.; Lei, J. T.; Wen, B.; Schnaubelt, M.; Krug, K.; Song, X.; Cieřlik, M.; Chang, H.-Y.; Wyczalkowski, M. A.; Li, K.; Colaprico, A.; Li, Q. K.; Clark, D. J.; Hu, Y.; Cao, L.; Pan, J.; Wang, Y.; Cho, K.-C.; Shi, Z.; Liao, Y.; Jiang, W.; Anurag, M.; Ji, J.; Yoo, S.; Zhou, D. C.; Liang, W.-W.; Wendl, M.; Vats, P.; Carr, S. A.; Mani, D. R.; Zhang, Z.; Qian, J.; Chen, X. S.; Pico, A. R.; Wang, P.; Chinnaiyan, A. M.; Ketchum, K. A.; Kinsinger, C. R.; Robles, A. I.; An, E.; Hiltke, T.; Mesri, M.; Thiagarajan, M.; Weaver, A. M.; Sikora, A. G.; Lubiński, J.; Wierzbicka, M.; Wiznerowicz, M.; Satpathy,

- S.; Gillette, M. A.; Miles, G.; Ellis, M. J.; Omenn, G. S.; Rodriguez, H.; Boja, E. S.; Dhanasekaran, S. M.; Ding, L.; Nesvizhskii, A. I.; El-Naggar, A. K.; Chan, D. W.; Zhang, H.; Zhang, B.; Clinical Proteomic Tumor Analysis Consortium. Proteogenomic Insights into the Biology and Treatment of HPV-Negative Head and Neck Squamous Cell Carcinoma. *Cancer Cell* **2021**, 39 (3), 361–379.e16. DOI: 10.1016/j.ccell.2020.12.007.
- (30) Gunning, P. W.; Ghoshdastider, U.; Whitaker, S.; Popp, D.; Robinson, R. C. The Evolution of Compositionally and Functionally Distinct Actin Filaments. *J. Cell Sci.* **2015**, 128 (11), 2009–2019. DOI: 10.1242/jcs.165563.
- (31) Hitchcock-DeGregori, S. E.; Barua, B. Tropomyosin Structure, Function, and Interactions: A Dynamic Regulator. *Subcell Biochem* **2017**, 82, 253–284. DOI: 10.1007/978-3-319-49674-0_9.
- (32) Geeves, M. A.; Hitchcock-DeGregori, S. E.; Gunning, P. W. A Systematic Nomenclature for Mammalian Tropomyosin Isoforms. *J. Muscle Res. Cell Motil.* **2015**, 36 (2), 147–153. DOI: 10.1007/s10974-014-9389-6.
- (33) Thierfelder, L.; Watkins, H.; MacRae, C.; Lamas, R.; McKenna, W.; Vosberg, H. P.; Seidman, J. G.; Seidman, C. E. Alpha-Tropomyosin and Cardiac Troponin T Mutations Cause Familial Hypertrophic Cardiomyopathy: A Disease of the Sarcomere. *Cell* **1994**, 77 (5), 701–712. DOI: 10.1016/0092-8674(94)90054-x.
- (34) Hershberger, R. E.; Norton, N.; Morales, A.; Li, D.; Siegfried, J. D.; Gonzalez-Quintana, J. Coding Sequence Rare Variants Identified in MYBPC3, MYH6, TPM1, TNNC1, and TNNT3 from 312 Patients with Familial or Idiopathic Dilated Cardiomyopathy. *Circ. Cardiovasc. Genet.* **2010**, 3 (2), 155–161. DOI: 10.1161/CIRCGENETICS.109.912345.
- (35) Fokstuen, S.; Munoz, A.; Melacini, P.; Iliceto, S.; Perrot, A.; Ozcelik, C.; Jeanrenaud, X.; Rieubland, C.; Farr, M.; Faber, L.; Sigwart, U.; Mach, F.; Lerch, R.; Antonarakis, S. E.; Blouin, J.-L. Rapid Detection of Genetic Variants in Hypertrophic Cardiomyopathy by Custom DNA Resequencing Array in Clinical Practice. *J. Med. Genet.* **2011**, 48 (8), 572–576. DOI: 10.1136/jmg.2010.083345.
- (36) Orzechowski, M.; Fischer, S.; Moore, J. R.; Lehman, W.; Farman, G. P. Energy Landscapes Reveal the Myopathic Effects of Tropomyosin Mutations. *Arch. Biochem. Biophys.* **2014**, 564, 89–99. DOI: 10.1016/j.abb.2014.09.007.
- (37) Olson, T. M.; Kishimoto, N. Y.; Whitby, F. G.; Michels, V. V. Mutations That Alter the Surface Charge of Alpha-Tropomyosin Are Associated with Dilated Cardiomyopathy. *J. Mol. Cell. Cardiol.* **2001**, 33 (4), 723–732. DOI: 10.1006/jmcc.2000.1339.
- (38) Barry, M. E.; Rynkiewicz, M. J.; Pavadai, E.; Viana, A.; Lehman, W.; Moore, J. R. Glutamate 139 of Tropomyosin Is Critical for Cardiac Thin Filament Blocked-State Stabilization. *J. Mol. Cell. Cardiol.* **2024**, 188, 30–37. DOI: 10.1016/j.yjmcc.2024.01.004.
- (39) F. Wiecek, D. The Role of Tropomyosin in Cardiac Function and Disease. In *Cardiac diseases and interventions in 21st century*; Karcioğlu, O., Ed.; IntechOpen, 2019. DOI: 10.5772/intechopen.81420.
- (40) Torre, R.; Freitas, M.; Costa-Rama, E.; Nouws, H. P. A.; Delerue-Matos, C. Food Allergen Control: Tropomyosin Analysis through Electrochemical Immunosensing. *Food Chem.* **2022**, 396, 133659. DOI: 10.1016/j.foodchem.2022.133659.

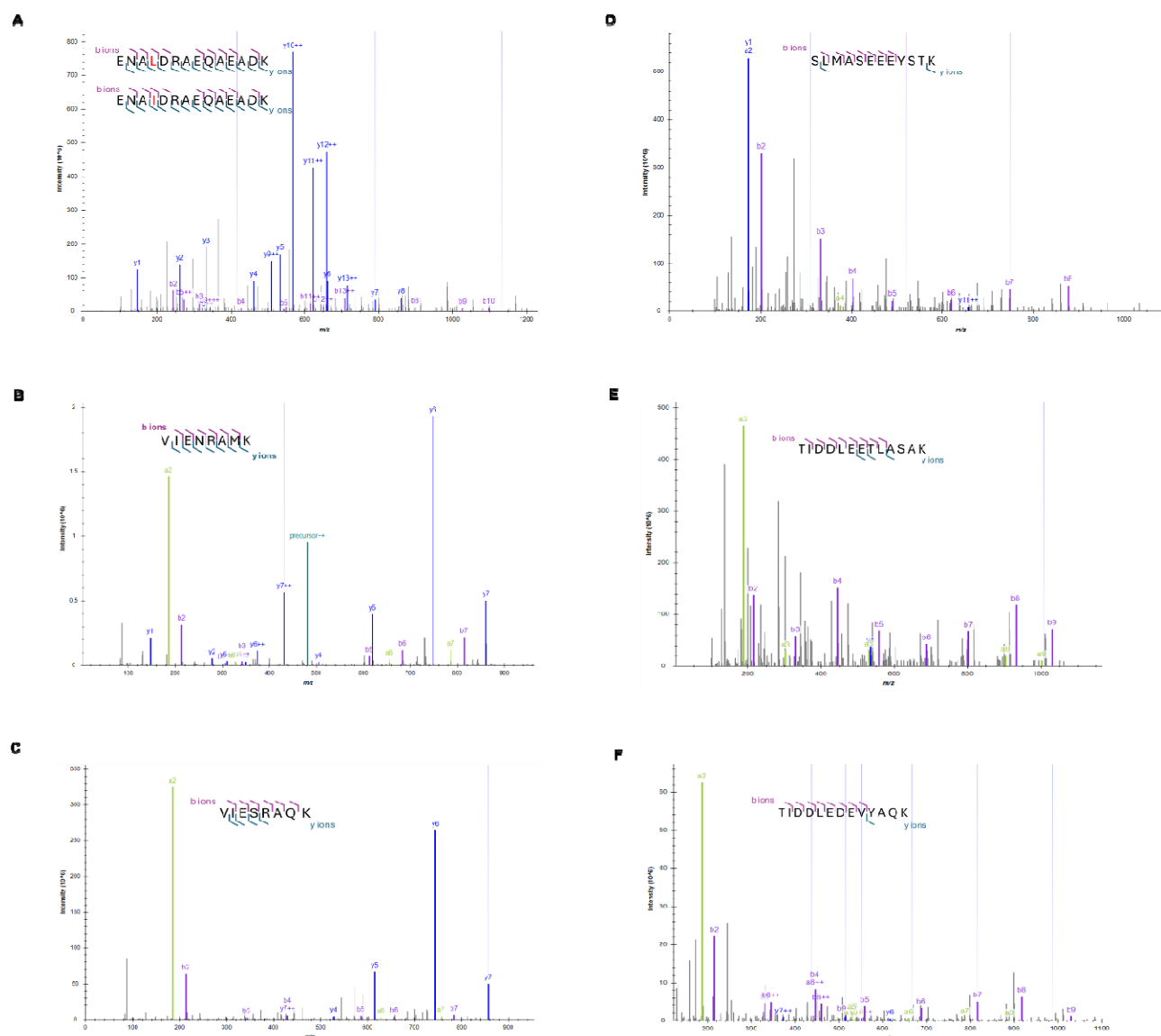
- (41) Abood, A.; Mesner, L. D.; Jeffery, E. D.; Murali, M.; Lehe, M.; Saquing, J.; Farber, C. R.; Sheynkman, G. M. Long-Read Proteogenomics to Connect Disease-Associated SQTs to the Protein Isoform Effectors of Disease. *BioRxiv* **2023**. DOI: 10.1101/2023.03.17.531557.
- (42) Jekel, P. A.; Weijer, W. J.; Beintema, J. J. Use of Endoproteinase Lys-C from *Lysobacter* Enzymogenes in Protein Sequence Analysis. *Anal. Biochem.* **1983**, *134* (2), 347–354.
- (43) Jewett, J. C.; Sletten, E. M.; Bertozzi, C. R. Rapid Cu-Free Click Chemistry with Readily Synthesized Biarylazacyclooctynones. *J. Am. Chem. Soc.* **2010**, *132* (11), 3688–3690. DOI: 10.1021/ja100014q.
- (44) Tang, H.-Y.; Beer, L. A.; Tanyi, J. L.; Zhang, R.; Liu, Q.; Speicher, D. W. Protein Isoform-Specific Validation Defines Multiple Chloride Intracellular Channel and Tropomyosin Isoforms as Serological Biomarkers of Ovarian Cancer. *J. Proteomics* **2013**, *89*, 165–178. DOI: 10.1016/j.jprot.2013.06.016.
- (45) Gerhard, D. S.; Wagner, L.; Feingold, E. A.; Shenmen, C. M.; Grouse, L. H.; Schuler, G.; Klein, S. L.; Old, S.; Rasooly, R.; Good, P.; Guyer, M.; Peck, A. M.; Derge, J. G.; Lipman, D.; Collins, F. S.; Jang, W.; Sherry, S.; Feolo, M.; Misquitta, L.; Lee, E.; Rotmistrovsky, K.; Greenhut, S. F.; Schaefer, C. F.; Buetow, K.; Bonner, T. I.; Haussler, D.; Kent, J.; Kiekhaus, M.; Furey, T.; Brent, M.; Prange, C.; Schreiber, K.; Shapiro, N.; Bhat, N. K.; Hopkins, R. F.; Hsie, F.; Driscoll, T.; Soares, M. B.; Casavant, T. L.; Scheetz, T. E.; Brownstein, M. J.; Usdin, T. B.; Toshiyuki, S.; Carninci, P.; Piao, Y.; Dudekula, D. B.; Ko, M. S. H.; Kawakami, K.; Suzuki, Y.; Sugano, S.; Gruber, C. E.; Smith, M. R.; Simmons, B.; Moore, T.; Waterman, R.; Johnson, S. L.; Ruan, Y.; Wei, C. L.; Mathavan, S.; Gunaratne, P. H.; Wu, J.; Garcia, A. M.; Hulyk, S. W.; Fuh, E.; Yuan, Y.; Sneed, A.; Kowis, C.; Hodgson, A.; Muzny, D. M.; McPherson, J.; Gibbs, R. A.; Fahey, J.; Helton, E.; Kettelman, M.; Madan, A.; Rodrigues, S.; Sanchez, A.; Whiting, M.; Madari, A.; Young, A. C.; Wetherby, K. D.; Granite, S. J.; Kwong, P. N.; Brinkley, C. P.; Pearson, R. L.; Bouffard, G. G.; Blakesly, R. W.; Green, E. D.; Dickson, M. C.; Rodriguez, A. C.; Grimwood, J.; Schmutz, J.; Myers, R. M.; Butterfield, Y. S. N.; Griffith, M.; Griffith, O. L.; Krzywinski, M. I.; Liao, N.; MGC Project Team. The Status, Quality, and Expansion of the NIH Full-Length CDNA Project: The Mammalian Gene Collection (MGC). *Genome Res.* **2004**, *14* (10B), 2121–2127. DOI: 10.1101/gr.2596504.
- (46) Prasad, G. L.; Meissner, S.; Sheer, D. G.; Cooper, H. L. A CDNA Encoding a Muscle-Type Tropomyosin Cloned from a Human Epithelial Cell Line: Identity with Human Fibroblast Tropomyosin TM1. *Biochem. Biophys. Res. Commun.* **1991**, *177* (3), 1068–1075. DOI: 10.1016/0006-291x(91)90647-p.
- (47) MacLeod, A. R.; Houlker, C.; Reinach, F. C.; Smillie, L. B.; Talbot, K.; Modi, G.; Walsh, F. S. A Muscle-Type Tropomyosin in Human Fibroblasts: Evidence for Expression by an Alternative RNA Splicing Mechanism. *Proc Natl Acad Sci USA* **1985**, *82* (23), 7835–7839. DOI: 10.1073/pnas.82.23.7835.
- (48) Liu, Y. A Peptidiform Based Proteomic Strategy for Studying Functions of Post-Translational Modifications. *Proteomics* **2022**, *22* (4), e2100316. DOI: 10.1002/pmic.202100316.
- (49) Hornbeck, P. V.; Kornhauser, J. M.; Tkachev, S.; Zhang, B.; Skrzypek, E.; Murray, B.; Latham, V.; Sullivan, M. PhosphoSitePlus: A Comprehensive Resource for Investigating the Structure and Function of Experimentally Determined Post-Translational

- Modifications in Man and Mouse. *Nucleic Acids Res.* **2012**, *40* (Database issue), D261-70. DOI: 10.1093/nar/gkr1122.
- (50) Salovska, B.; Zhu, H.; Gandhi, T.; Frank, M.; Li, W.; Rosenberger, G.; Wu, C.; Germain, P.-L.; Zhou, H.; Hodny, Z.; Reiter, L.; Liu, Y. Isoform-Resolved Correlation Analysis between mRNA Abundance Regulation and Protein Level Degradation. *Mol. Syst. Biol.* **2020**, *16* (3), e9170. DOI: 10.15252/msb.20199170.
- (51) Stokes, M. P.; Farnsworth, C. L.; Moritz, A.; Silva, J. C.; Jia, X.; Lee, K. A.; Guo, A.; Polakiewicz, R. D.; Comb, M. J. PTMScan Direct: Identification and Quantification of Peptides from Critical Signaling Proteins by Immunoaffinity Enrichment Coupled with LC-MS/MS. *Mol. Cell. Proteomics* **2012**, *11* (5), 187–201. DOI: 10.1074/mcp.M111.015883.
- (52) Blackburn, K.; Goshe, M. B. Challenges and Strategies for Targeted Phosphorylation Site Identification and Quantification Using Mass Spectrometry Analysis. *Brief. Funct. Genomic. Proteomic.* **2009**, *8* (2), 90–103. DOI: 10.1093/bfgp/eln051.
- (53) Dube, D. K.; Dube, S.; Abbott, L.; Wang, J.; Fan, Y.; Alshiekh-Nasany, R.; Shah, K. K.; Rudloff, A. P.; Poiesz, B. J.; Sanger, J. M.; Sanger, J. W. Identification, Characterization, and Expression of Sarcomeric Tropomyosin Isoforms in Zebrafish. *Cytoskeleton (Hoboken)* **2017**, *74* (3), 125–142. DOI: 10.1002/cm.21352.
- (54) James, J. K.; Pike, D. H.; Khan, I. J.; Nanda, V. Structural and Dynamic Properties of Allergen and Non-Allergen Forms of Tropomyosin. *Structure* **2018**, *26* (7), 997-1006.e5. DOI: 10.1016/j.str.2018.05.002.
- (55) Sicherer, S. H.; Muñoz-Furlong, A.; Sampson, H. A. Prevalence of Seafood Allergy in the United States Determined by a Random Telephone Survey. *J. Allergy Clin. Immunol.* **2004**, *114* (1), 159–165. DOI: 10.1016/j.jaci.2004.04.018.
- (56) Marcu, A.; Bichmann, L.; Kuchenbecker, L.; Kowalewski, D. J.; Freudenmann, L. K.; Backert, L.; Mühlenbruch, L.; Szolek, A.; Lübke, M.; Wagner, P.; Engler, T.; Matovina, S.; Wang, J.; Hauri-Hohl, M.; Martin, R.; Kapolou, K.; Walz, J. S.; Velz, J.; Moch, H.; Regli, L.; Silginer, M.; Weller, M.; Löffler, M. W.; Erhard, F.; Schlosser, A.; Kohlbacher, O.; Stevanović, S.; Rammensee, H.-G.; Neidert, M. C. HLA Ligand Atlas: A Benign Reference of HLA-Presented Peptides to Improve T-Cell-Based Cancer Immunotherapy. *J. Immunother. Cancer* **2021**, *9* (4). DOI: 10.1136/jitc-2020-002071.
- (57) Chong, C.; Müller, M.; Pak, H.; Harnett, D.; Huber, F.; Grun, D.; Leleu, M.; Auger, A.; Arnaud, M.; Stevenson, B. J.; Michaux, J.; Bilic, I.; Hirsekorn, A.; Calviello, L.; Simó-Riudalbas, L.; Planet, E.; Lubiński, J.; Bryśkiewicz, M.; Wiznerowicz, M.; Xenarios, I.; Zhang, L.; Trono, D.; Harari, A.; Ohler, U.; Coukos, G.; Bassani-Sternberg, M. Integrated Proteogenomic Deep Sequencing and Analytics Accurately Identify Non-Canonical Peptides in Tumor Immuno-peptidomes. *Nat. Commun.* **2020**, *11* (1), 1293. DOI: 10.1038/s41467-020-14968-9.

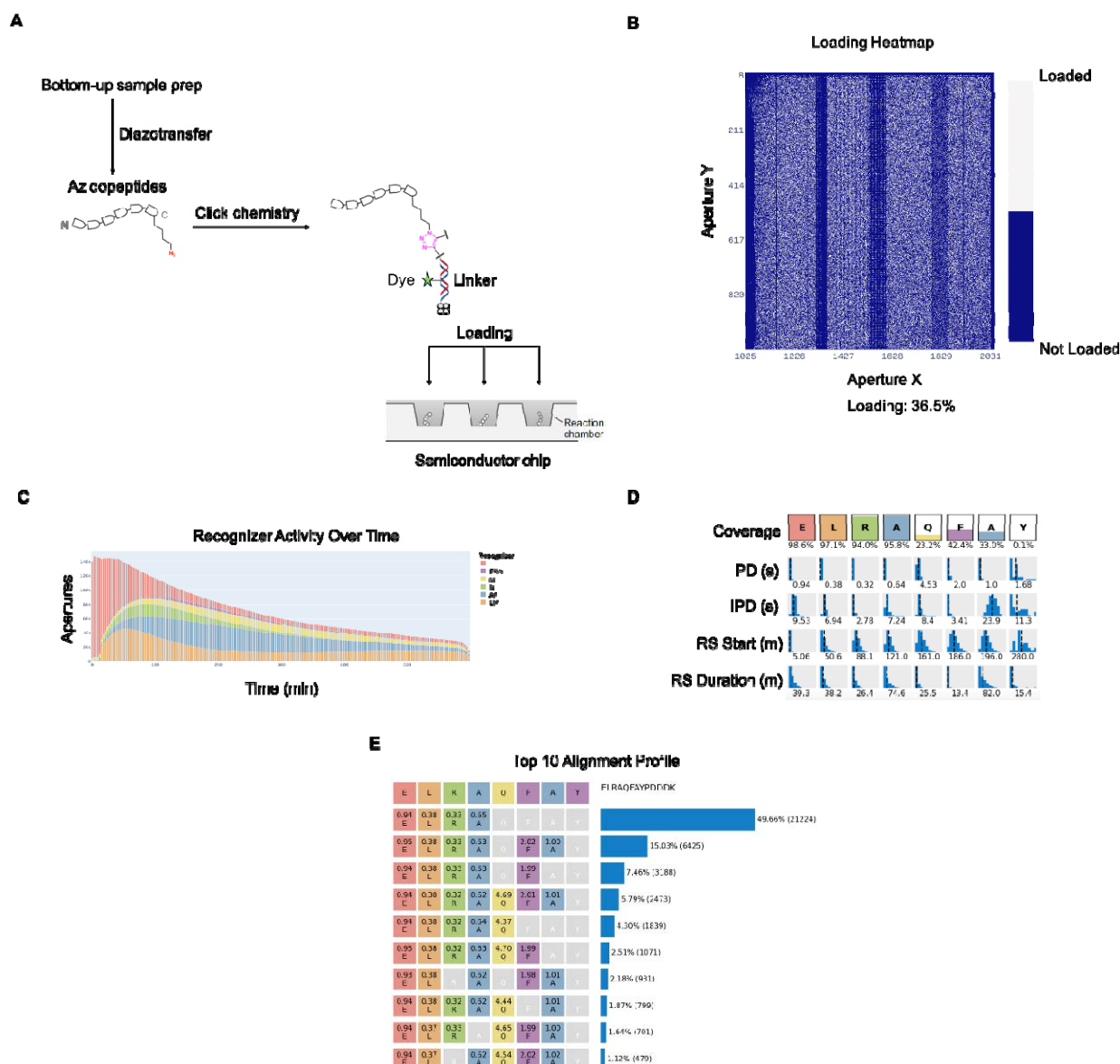
Supplementary data



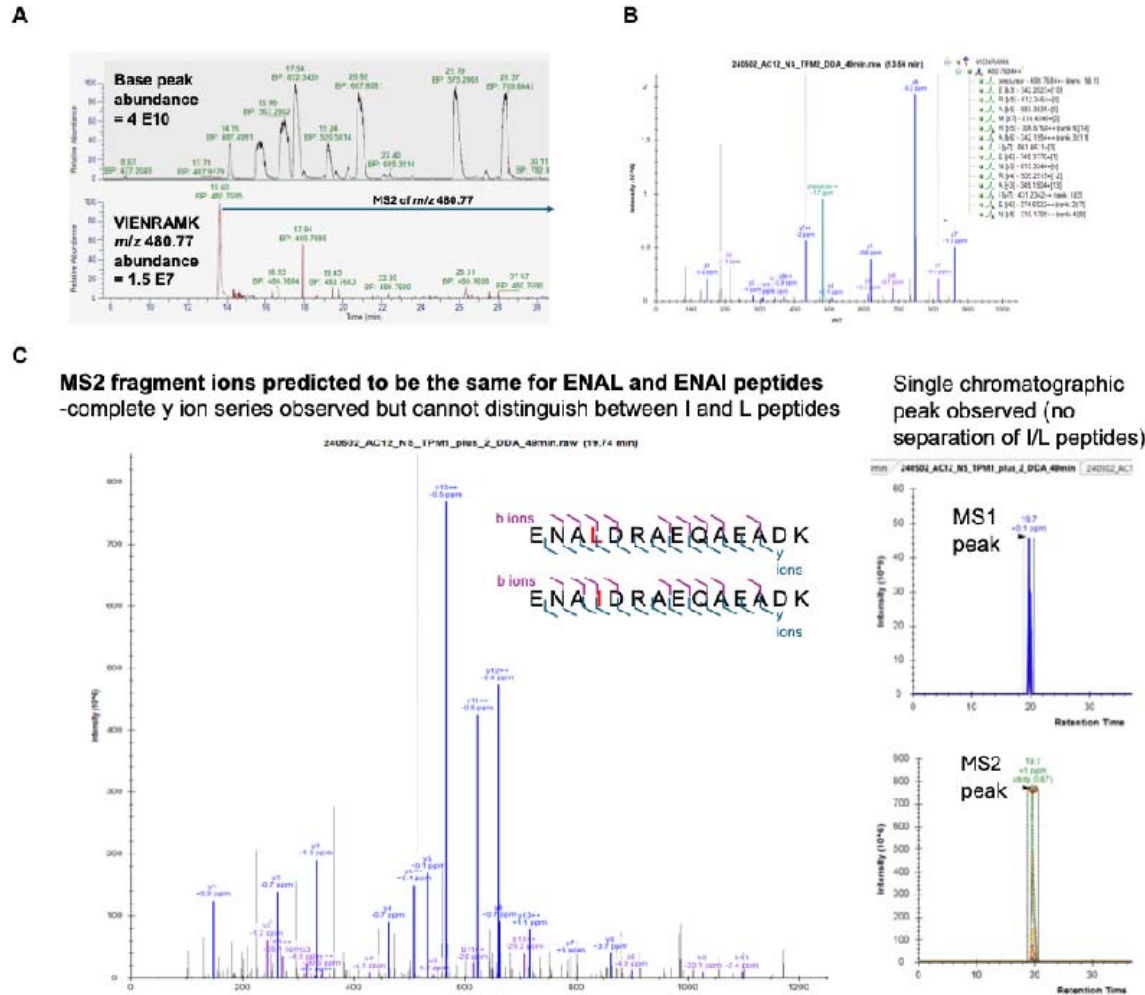
Supplementary Figure 1. Exonic structure of the human tropomyosin genes. Most vertebrates have four tropomyosin genes, each encoding multiple isoforms due to alternate exon expression. Each tropomyosin gene is composed of 11 to 15 exons, with different splicing patterns correlated with different functions. The *white* boxes are constitutively expressed in all spliceforms of the gene, while *colored* boxes are alternatively spliced to produce distinct spliceforms. The most variable regions of tropomyosin are the ends, encoded by alternate start and end exons that form the intermolecular overlap junction.



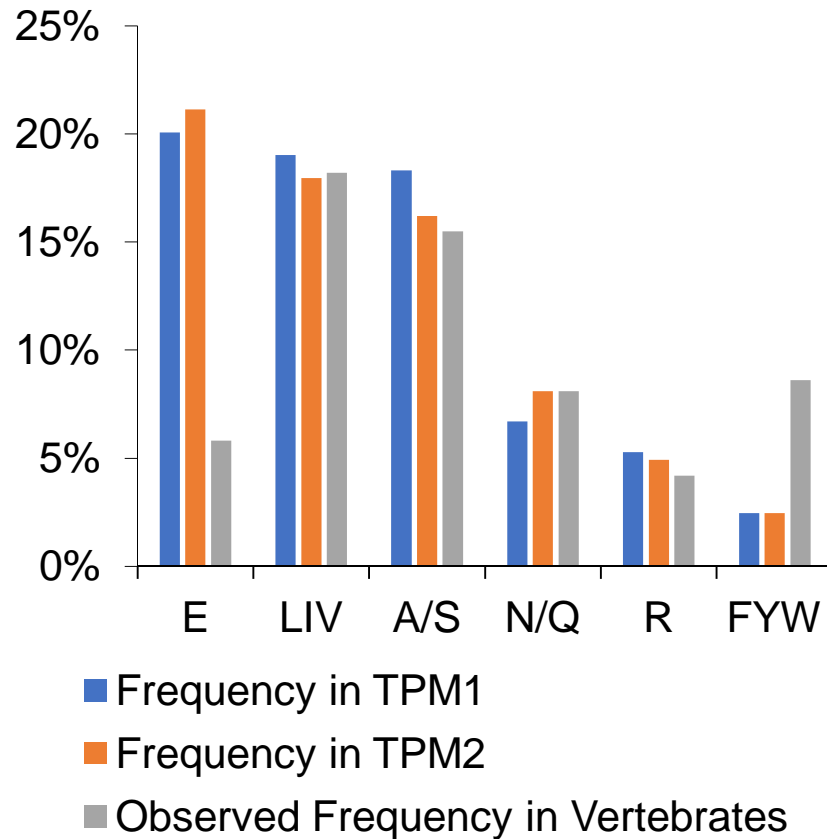
Supplementary Figure 2. MS2 spectra and fragmentation map of 7 synthetic TPM peptides. A-F shows the ion intensities and fragmentation of each peptide: (A) ENA[L/I]DRAEQAEADK (spectra are indistinguishable from each other), (B) VIENRAMK, (C) VIESRAQK, (D) SLMASEEEYSTK, (E) TIDDLEETLASAK, (F) TIDDLEDEVYAQK VIESRAQK, (D) SLMASEEEYSTK, (E) TIDDLEETLASAK, (F) TIDDLEDEVYAQK



Supplementary Figure 3. Schematic of Platinum library preparation, chip loading, and peptide alignment. (A) Peptides are subjected to diazotransfer to install an N₃ group at the side chain of the C-terminal lysine. The derivatized peptides are conjugated to fluorescently labeled linkers via click chemistry, then loaded onto flow cells on a semiconductor chip and immobilized in nanoscale reaction chambers (or apertures). (B) A heatmap of aperture occupancy is generated after on-chip peptide loading. White pixels represent loaded apertures, while dark blue pixels are unoccupied apertures. (C) Recognizer activity over the course of a sequencing experiment. Number of apertures registering pulsing activity for each recognizer type are recorded and plotted in real-time. (D) Example coverage map of a synthetic peptide ELRAQFAYPDDK. Colored boxes show the frequency of detection for each residue. The distribution and median of each of the kinetic parameters collected by Platinum are displayed as blue histograms underneath each residue. (E) Reads are grouped into different alignment profiles based on the RSs they exhibit. The number of reads in each alignment profile for this peptide are displayed as blue bars to the left of the profiles.



Supplementary Figure 4. Recombinant TPM2 subjected to endoproteinase Lys-C and LC-MS/MS (Orbitrap Eclipse). (A) VIENRAMK peptide ion abundance is 1000-fold less than base peak abundance, leading to a single MS2 scan acquired during DDA method. **(B)** A spectrum of VIENRAMK with b/y ion series. **(C)** Single chromatographic peak observed with isobaric peptides ENALDRAEQAEADK (TPM1) and ENAI DRAEQAEADK (TPM2) complicates LC-MS/MS detection of paralog-specific peptides.



Supplementary Figure 5. TPM1/2 amino acid frequency of NAA recognizers compatible with Platinum.

New Results

Follow this preprint Previous

Next

Protein sequencing with single amino acid resolution discerns peptides that discriminate tropomyosin proteoforms

Natchanon Sittipongpittaya, Kenneth Skinner, Erin Jeffery, Emily Watts-Whitehead, Gloria M Sheynkman

doi: <https://doi.org/10.1101/2024.11.04.621980>

This article is a preprint and has not been certified by peer review [what does this mean?].



Abstract Info/History Metrics Preview PDF

Abstract

Protein variants of the same gene--proteoforms--can have high molecular similarity yet exhibit different biological functions. Thus, identifying unique peptides that unambiguously map to proteoforms can provide crucial biological insights. In humans, four human tropomyosin (TPM) genes produce similar proteoforms that can be challenging to distinguish with standard proteomics tools. For example, TPM1 and TPM2 share 85% sequence identity, with amino acid substitutions that play unique roles in muscle contraction and myopathies. In this study, we evaluated the ability of the recently released Platinum single-molecule protein sequencer to detect proteoform-informative peptides. Platinum employs fluorophore-labeled recognizers that reversibly bind to cognate N-terminal amino acids (NAAs), enabling polypeptide sequencing within nanoscale apertures of a semiconductor chip that can accommodate single peptide molecules. As a proof of concept, we evaluated the ability of Platinum to distinguish three main types of proteoform variation: paralog-level, transcript-level, and post-translational modification (PTM). We distinguished paralogous TPM1 and TPM2 peptides differing by a single isobaric residue (leucine/isoleucine). We also distinguished tissue-specific TPM2 spliceforms. Notably, we found that a phosphotyrosine-modified peptide displayed reduced recognizer affinity for tyrosine, showing sensitivity to PTMs. This study paves the way for the targeted detection of proteoform biomarkers at the single molecule level

Competing Interest Statement

KAS is an employee and shareholder of Quantum-Si. No other authors have any disclosures.

Copyright The copyright holder for this preprint is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-NC-ND 4.0 International license](#).

bioRxiv and medRxiv thank the following for their generous financial support:

The Chan Zuckerberg Initiative, Cold Spring Harbor Laboratory, the Sergey Brin Family Foundation, California Institute of Technology, Centre National de la Recherche Scientifique, Fred Hutchinson Cancer Center, Imperial College London, Massachusetts Institute of Technology, Stanford University, University of Washington, and Vrije Universiteit Amsterdam.

Back to top

Posted November 06, 2024.

Download PDF
 Print/Save Options

Email
 Share
 Citation Tools
 Get QR code

Post Like 0

COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv

Subject Area

Genomics

Subject Areas

All Articles

Animal Behavior and Cognition
Biochemistry
Bioengineering
Bioinformatics
Biophysics
Cancer Biology
Cell Biology
Clinical Trials*
Developmental Biology
Ecology
Epidemiology*
Evolutionary Biology
Genetics
Genomics
Immunology
Microbiology
Molecular Biology
Neuroscience
Paleontology
Pathology
Pharmacology and Toxicology
Physiology
Plant Biology
Scientific Communication and Education
Synthetic Biology
Systems Biology
Zoology

* The Clinical Trials and Epidemiology subject categories are now closed to new submissions following the completion of bioRxiv's clinical research pilot project and launch of the dedicated health sciences server medRxiv (submit.medrxiv.org). New papers that report results of Clinical Trials must now be submitted to medRxiv. Most new Epidemiology papers also should be submitted to medRxiv, but if a paper contains no health-related information, authors may choose to submit it to another bioRxiv subject category (e.g., Genetics or Microbiology).