Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps

Tiannan Guo¹, Petri Kouvonen¹,², Ching Chiek Koh¹,³, Ludovic C Gillet¹, Witold E Wolski¹,⁴, Hannes L Röst¹,⁵, George Rosenberger¹,⁵, Ben C Collins¹, Lorenz C Blum⁴,⁶, Silke Gillessen⁷, Markus Joerger⁷, Wolfram Jochum⁸ & Ruedi Aebersold¹,⁹

Clinical specimens are each inherently unique, limited and nonrenewable. Small samples such as tissue biopsies are often completely consumed after a limited number of analyses. Here we present a method that enables fast and reproducible conversion of a small amount of tissue (approximating the quantity obtained by a biopsy) into a single, permanent digital file representing the mass spectrometry (MS)-measurable proteome of the sample. The method combines pressure cycling technology (PCT) and sequential window acquisition of all theoretical fragment ion spectra (SWATH)-MS. The resulting proteome maps can be analyzed, re-analyzed, compared and mined in silico to detect and quantify specific proteins across multiple samples. We used this method to process and convert 18 biopsy samples from nine patients with renal cell carcinoma into SWATH-MS fragment ion maps. From these proteome maps we detected and quantified more than 2,000 proteins with a high degree of reproducibility across all samples. The measured proteins clearly distinguished tumorous kidney tissues from healthy tissues and differentiated distinct histomorphological kidney cancer subtypes.

For many diseases, especially cancers, examination of disease-affected tissues is essential for clinical diagnosis, prognosis and therapy selection. Traditionally, tissues are fixed in formalin, embedded in paraffin and examined empirically (by light microscopy, for example) by highly trained pathologists; the detected histomorphological features are then used to diagnose and classify diseases.

More recently, the use of molecular patterns to support histomorphological classification of tissues has attracted interest¹. mRNA transcript profiling has been used to group breast cancer tissues into classes based on predicted clinical course²,³; analogous strategies have been applied to most other common cancer types⁴. Spectral detection arrays in which specific proteins are detected and quantified via arrays of reverse arrays in which specific proteins are detected and quantified via arrays of available antibodies in extracts from tissue specimens¹⁰. Although these methods increase the degree of multiplexing and thus support, in principle, the analysis of protein patterns, they remain limited by the repertoire of available reagents, their quantitative accuracy and their dynamic range.

Proteins are arguably the most informative class of biomolecules for tissue classification, not only because they most directly indicate the functional state of disease-relevant molecular systems such as signaling and metabolic pathways, but also because they are themselves the major drug targets in many diseases, particularly cancers. Antibody-based assays have been widely used for protein measurement in medical research. For tissue classification, elaborate immunohistochemical staining protocols have been developed and are widely applied. However, these analytical tools are restricted by the paucity and high cost of validated reagents, difficult batch-to-batch quality control and cross-reproducibility issues. Moreover, the limited ability to multiplex these analyses precludes the measurement of multiple protein panels in biopsy tissue samples. Many array-based methods have been developed to alleviate some of these limitations. These include tissue microarrays⁹ in which a tissue sample is cut into multiple sections, each of which is tested by a specific immunostain, and reverse arrays in which specific proteins are detected and quantified via arrays of available antibodies in extracts from tissue specimens¹⁰. Although these methods increase the degree of multiplexing and thus support, in principle, the analysis of protein patterns, they remain limited by the repertoire of available reagents, their quantitative accuracy and their dynamic range.

MS has evolved as the method of choice for versatile, multiplexed and quantitative analysis of proteins and proteomes. So far, two main MS strategies have been used for protein analysis, each of which is represented by a multiplicity of methods. Both strategies depend on the efficient and reproducible extraction of proteins from the biological source and the transformation of the protein sample into a complex peptide sample by proteolysis. In both strategies, the peptide sample is then analyzed by MS. The first strategy, referred to as ‘discovery’ or ‘shotgun’ proteomics, is characterized by data-dependent acquisition (DDA). DDA uses a simple heuristic to instruct the mass spectrometer to sequentially select peptide ions from the multitude of precursor ions detected in a survey scan, to fragment the isolated...
ions and to record their fragment ion spectra. This strategy is capable of identifying thousands of peptides per sample. However, if the number of peptides in the sample substantially exceeds the number of sequencing cycles available during data acquisition—as is the case for unfractionated tissue extracts—the population of correctly identified peptides across multiple samples becomes poorly reproducible. The second strategy, referred to as ‘targeted’ proteomics, is exemplified by selected/multiple reaction monitoring (S/MRM). In targeted proteomics, predetermined sets of peptides are selected from the available sample peptides and identified and quantified in a sample by means of specific MS assays that have to be prepared a priori for each targeted peptide. S/MRM generates data sets that are highly reproducible, quantitatively accurate and sensitive, but its capacity to multiplex is limited to ~100 proteins per analysis. Multiplexing up to a few hundred proteins using S/MRM is possible, albeit generally at the cost of diminished quantification accuracy or limit of detection.

We recently introduced an MS strategy that allows reproducible detection and accurate quantification of thousands of proteins from complex protein samples with performance characteristics that approach those of S/MRM. The method, SWATH-MS\(^1\), consists of data-independent acquisition (DIA), which generates fragment ion spectra from all MS-measurable peptides of a proteome within a user-defined retention time versus mass-to-charge ratio space, as well as a post-acquisition targeted data analysis strategy that associates signature fragment ion patterns for targeted peptides with feature groups in the SWATH-MS signal map, thus confirming the presence and determining the relative quantity of query peptides among samples. As the SWATH-MS data sets comprise a single, permanent digital file representing the gap-free, MS-measurable proteome of the sample, once acquired, they can be perpetually analyzed \textit{in silico} to test new hypotheses (for example, to quantify newly postulated measurable biomarkers across suitable sample cohorts). At a sufficient degree of pattern reproducibility, the method also supports cross-study comparisons that are not anticipated at the time of data acquisition. To date, SWATH-MS has been used to study the relatively uncomplicated yeast proteome\(^2\), the subproteome of glycosylated proteins\(^3\) and purified protein complexes\(^4\), but not to a proteome as complex as that of a human tissue.

Regardless of the MS strategy used for tissue proteome analysis, the success of the measurements critically depends on the quality of the analyzed peptide sample. Numerous published protocols notwithstanding, reproducible generation of high yields of peptide samples from small amounts of tissue, such as those from biopsies, remains challenging. PCT recently emerged as a technology that can be used to perform various steps in proteomic sample preparation\(^5\).

In brief, programmed cycles oscillating between ultrahigh and ambient pressure induce dissolution of matrix and proteins, facilitating and accelerating proteolysis\(^6\). For clinical samples, PCT also inactivates potential contaminant infectious microorganisms\(^7\).

In this report, we describe a PCT-based protocol for MS-ready peptide preparation from biopsy-level tissues. The protocol integrates tissue lysis with protein extraction and digestion and, in combination with SWATH-MS, it rapidly generates highly reproducible fragment ion maps from human tissue biopsies. This method may enable many applications in basic and clinical research as well as in digital biobanking.

RESULTS

Establishment of the PCT-SWATH workflow

To start, we optimized a PCT-based protocol that integrated tissue lysis and protein extraction and digestion (Fig. 1). Briefly, we placed a small piece of wet biopsy tissue (about 1 mg) into a pressure-resistant MicroTube, added protein extraction and digestion reagents, and subjected the mixture to pressure cycling. Our method was different from existing methods of tissue sample preparation for proteomic analysis in that the entire procedure—comprising tissue lysis, protein extraction and protein digestion—was integrated into a seamless protocol with minimal sample transfer. This permitted analysis of a minimal sample amount as well as precise control (by Barocycler) of the reaction solvent volume and incubation time for each step, effectively minimizing technical variation. We cleaned the resulting peptides on reversed-phase columns, after which they were ready for SWATH-MS.

Protein extraction and digestion were optimized with respect to yield, reproducibility and compatibility of the samples with MS. Among the many tissue lysis and protein extraction conditions tested, the combination of PCT and protein digestion protocol with minimal sample transfer was the best. The peptides were then analyzed using SWATH-MS, which includes target ion maps from human tissue biopsies. This method may enable many applications in basic and clinical research as well as in digital biobanking.

Figure 1 PCT-SWATH method flow chart. (a) Tissue lysis and protein extraction. A batch of six biopsy-scale tissues are placed in MicroTubes and lysed in the Barocycler for 60 min. (b) Protein digestion. Extracted proteins are reduced and alkylated. The urea concentration is then diluted before Lys-C digestion in the Barocycler for 45 min. The MicroTubes are then taken out of the Barocycler, and the urea concentration is further diluted before trypsin digestion for 90 min under Barocycling. The resultant peptides are desalted using C18 cartridges and dried under vacuum. This process can be completed in about 6 h. (c) The peptide samples are then analyzed using SWATH-MS, which takes about 3 hours. Each SWATH cycle contains an MS1 scan of all peptide ions and 32 sequential MS2 scans of peptide fragment ions from peptide ions in a 25-amu-wide window. The scheme of the SWATH window is described in the Online Methods. (d) The data are analyzed using OpenSWATH and a SWATH assay library. Each peptide fragment ion signature is scored based on its similarity to the respective reference signature in the SWATH assay library. Decoy assays are also scored to calculate the false discovery rate. Feature alignment is performed for the confidently identified peptides across different samples.
a mixture of 8 M urea and 100 mM ammonium bicarbonate supplemented with a protease inhibitor cocktail, in combination with a cycling program of 50 s of ultrahigh pressure (45,000 p.s.i.) and 10 s of ambient pressure (14.7 p.s.i.), was found most efficient. Sixty of these pressure cycles (corresponding to a 60-min run time) were sufficient to extract proteins from the different sources tested, including human kidney tissue, the osteosarcoma cell line U2OS and mouse liver.

Protein digestion was also accelerated by PCT performed at 33 °C in the dark. Under this condition, the reducing agent tris(2-carboxyethyl)phosphine and the alkylation agent iodoacetamide were stable, allowing protein reduction and alkylation to occur in a single step (at yields that were equivalent to conventional methods that use a glass homogenizer for tissue lysis and protein extraction, perform sequential reduction and alkylation, and finish with overnight digestion in solution) (Fig. 2). Sequential Lys-C and trypsin proteolysis steps were used to efficiently digest proteins into peptides17. Digestion was accelerated using PCT with a cycling scheme of 50 s at 20,000 p.s.i. and 10 s at ambient pressure. Lys-C digestion was carried out in 6 M urea for 45 cycles, whereas trypsin digestion was performed in 1.6 M urea for 90 cycles. The entire process from tissue to proteolyzed, injection-ready peptide samples took 6–8 h. Six samples were processed concurrently in a single batch using this protocol (Fig. 1). A higher number of samples can be processed simultaneously with newer models of the Barocycler.

To assess the yield and reproducibility of the optimized PCT method, we determined the total peptide amount collected from a human cell line (osteosarcoma cell line U2OS) and two types of tissues (mouse liver and human kidney). In every case, the total amount of processed, MS injection-ready peptide was determined based on its absorbance at 280 nm using a NanoDrop spectrophotometer. From U2OS cells, approximately 116 µg of total peptide mass was generated per million cells (Fig. 2a). From aliquots of 1–3 mg samples of wet mouse liver and human kidney tissue, the method produced about 50 µg of total peptide mass per milligram of tissue with a high degree of reproducibility for technical replicates (coefficient of variation (CV) < 8%). The CV of peptide yield increased to 41% for 12 human kidney tissues from different patients, probably because of biological variation.

It was difficult to compare the yield of the PCT-assisted lysis and digestion method with that achieved by conventional methods on the same scale because we found it impractical to efficiently lyse and digest such small tissue samples using conventional glass homogenization or bead mill-based methods. Therefore, to compare the yield of conventional and PCT assisted methods, we lysed about 100 mg of mouse liver using a glass homogenizer. From the resulting lysate, we subjected 1% to conventional overnight in-solution digestion using Lys-C and trypsin sequentially. The peptide yield of this conventional method was about 30 µg per milligram tissue, 40% lower than the median yield of the PCT-based method carried out entirely on 1 mg (wet weight) of total tissue. We analyzed the respective peptide samples by the widely used DDA method in an ion-trap mass spectrometer and identified peptides by protein sequence database searching. The PCT method produced peptides with a similar length and charge state distribution compared with the conventional method, and the number of peptides identified was 10–20% higher for PCT (Fig. 2b).

Generation of SWATH-MS maps

We next used SWATH-MS to acquire fragment ion spectra of the peptide samples generated by PCT. The peptides were separated by an optimized reversed-phase gradient of 135 min, ionized by electrospray ionization and injected into a TripleTOF 5600 MS instrument operated in SWATH acquisition mode. Precursors with mass-to-charge ratios in the range of 400–1,200 were divided into 32 SWATH windows12. We fragmented all ionized peptide precursors in this mass range, generating comprehensive fragment ion maps. Identification of peptides from the SWATH map requires a priori information in the form of a SWATH assay library containing characteristic peptide fragment ion patterns and retention time information for all targeted peptides. We conducted targeted data analysis using the OpenSWATH software tool18.

We evaluated the reproducibility of the PCT-SWATH workflow by digesting four aliquots of a human kidney tissue sample and producing SWATH maps in triplicate. The 12 SWATH-MS maps were analyzed using OpenSWATH and a SWATH assay library containing 49,959 reference spectra for 41,542 proteotypic peptides from 4,624 reviewed Swiss-Prot proteins, compiled using DDA analyses of the kidney tissues in a TripleTOF 5600 mass spectrometer. Only one protein product from each gene was included in this analysis; therefore, we excluded protein isoforms. We performed feature alignment for peptides identified in different samples and protein-level quantification using the aLFQR R package19. At an estimated false discovery rate of 0.1% at precursor (ionized peptide) level, we quantified 1,632 unique proteins (Supplementary Table 1). The overall sparsity of the quantified matrix across the 1,632 proteins and 12 samples was 12%, as visualized in a heat map, indicating a high degree of data completeness for the method (Supplementary Fig. 1a).

Next we assessed variation among technical triplicates for each of the four digests. We determined the CV of triplicate injections of
the fourdigests in three batches and plotted the CV distributions as violin plots (Fig. 3a). Peptides not quantified in all samples were not included in the CV analysis. The median CV for repeated MS injection batches was between 10% and 12%. The variation of PCT-assisted digestion across the samples was slightly higher, with the median rising to 16–20% (Fig. 3a). The median CV of all 12 replicates was 24% (Fig. 3b). This is the overall median technical variation expected for the same tissue sample measured using the entire PCT-SWATH workflow. This value is comparable to that achieved in our laboratory by S/MRM in a label-free quantification mode when applied at high throughput to the model organism *Saccharomyces cerevisiae*.

The CV values were not dependent on the abundance of proteins (Supplementary Fig. 2).

### Dynamic range of proteins identified in kidney tissues

We applied the PCT-SWATH method to study a panel of 18 kidney tissue samples from nine patients with renal cell carcinoma (RCC). From each individual, two biopsies were collected: one tumorous and the other nontumorous, as determined by histomorphological examination. Among the nine patients, six had clear cell RCC (ccRCC), two had papillary RCC (pRCC) and one had chromophobe RCC (chRCC) (Supplementary Table 2). Each sample was measured in technical duplicates. The resulting SWATH-MS maps were searched against a kidney tissue SWATH assay library. In total, 2,375 proteins were quantified across all samples with a precursor FDR below 0.1% and a sparsity of 32% (Supplementary Table 3 and Supplementary Fig. 1b). The sparsity of the data matrix was higher than the previously tested 12 replicates, probably because of increased biological heterogeneity between individuals and between tumor and nontumor tissues.

Next we searched the literature and identified 53 Swiss-Prot proteins presently used as diagnostic or prognostic biomarkers for renal cancers, of which 21, including the widely used biomarkers VIM and AMACR, were quantified in our data set (Fig. 4). As we used exclusively proteotypic peptides, that is, peptides that are unique to a specific protein, we were able to independently quantify even highly concordant protein isoforms. This is exemplified by cytokeratin isoforms, some of which showed idio typic quantitative patterns across samples: after unsupervised clustering, most cytokeratins clustered together but cytokeratin 8 displayed a unique pattern (Fig. 4).

Next we examined the dynamic range of proteins quantified in kidney tissues on the basis of their signal intensities and compared these data to a few published data sets that contain estimates of absolute protein abundance in human cell lines. The dynamic range was about four orders of magnitude (Supplementary Fig. 3a), which is in line with the dynamic range estimations determined on standardized samples with the SWATH method. DDA analysis of extensively fractionated cell line or tissue lysates have reached a dynamic range of 4.6–5.8 orders of magnitude (Supplementary Fig. 3a).

To understand the subcellular localization of kidney proteins analyzed by PCT-SWATH, we analyzed the Gene Ontology annotations of the proteins identified in our data sets, the reference data sets as well as the entire human proteome. More mitochondrial proteins and fewer than nuclear proteins were present in our SWATH data compared to the reference populations (Supplementary Fig. 3b), indicating that low-abundance proteins were under-represented in the SWATH data. We also observed that the percentage of surface proteins in our SWATH data was slightly higher than than the DDA data from cell lines and tissues, consistent with a study showing that PCT-assisted lysis and digestion facilitated the detection of membrane proteins (Supplementary Fig. 3b). It is not surprising that the percentage of proteins from extracellular regions was high in both...
Classical proteome analysis of human kidney tissue proteomes by SWATH-MS. The violin plots show the distribution of CV values from injection replicates (n = 18), within-patient proteomic variation based on all RCC (n = 36) and ccRCC (n = 24) tissues, and among-patient variation based on all RCC (n = 262,144 for all, n = 512 for normal, n = 512 for tumor) and ccRCC (n = 4,096 for all, n = 64 for normal, n = 64 for tumor) tissues. The violin plot is as defined in the legend for Figure 3.

Variation analysis of tissue proteome PCT-SWATH maps

The availability of PCT-SWATH maps of the nine paired kidney biopsies allowed us to again determine the technical and biological variation of PCT-SWATH analysis. The median CV of the injection batch variation was below 6%, consistent with the 12 kidney tissue test replicates data discussed above (Fig. 5). To estimate biological variation within an individual, we calculated the median CV of all proteins in the SWATH data quantified in every pair of tissue biopsies, either nontumorous or tumorous, from each of the nine RCC patients. The median of median CV values that indicate within-patient variation rose to close to 40% (Fig. 5). Inter-patient CV values in paired kidney biopsies were slightly higher than intra-patient CV values (Fig. 5). Notably, the median CV value of nontumorous tissues from patients with ccRCC was 19%, lower than the CV observed in all other variations except injection batch variation (Fig. 5). The tumor tissues among patients, even in the ccRCC subtype, displayed relatively high variation (Fig. 5).

Classification of kidney tumor biopsies using PCT-SWATH

Reproducible quantification of clinically relevant proteins as outlined above shows high potential as a method for tumor classification based on proteome profiling. We found that the proteomic profile is useful for categorizing tissue samples according to their degree of malignancy and histomorphological subtypes (Fig. 6). Unsupervised clustering of more than 2,000 consistently quantified proteins resulted in the distinction of tumor biopsies from neighboring nontumorous biopsies from ccRCC patients (Fig. 6a). In total, 317 proteins were downregulated and 296 proteins were upregulated in tumor compared to nontumorous biopsies (Supplementary Table 4). Affected proteins included protein kinases and transcription factors, in addition to other proteins involved in the regulation of biological processes or pathways including apoptosis, metabolic pathways, signaling and immune response (Fig. 6c).

Unsupervised clustering also correctly distinguished ccRCC tumors from pRCC tumors (Fig. 6b). AMACR is used in clinic as a protein marker to distinguish pRCC (positive by immunohistochemical analysis) and ccRCC (negative by immunohistochemical analysis)21. Here the SWATH signal of AMACR was more than 13 times higher in pRCC than in ccRCC (P > 0.000001). GSTA1 and VIM are known to be highly expressed proteins in ccRCC compared to the other subtypes21, and our data was consistent with this (Supplementary Table 5).
**DISCUSSION**

Here we introduce a fairly simple and fast methodology to perform high-throughput, quantitative and highly reproducible proteomic analyses of clinical tissue biopsies. PCT-SWATH requires less than 12 h (from tissue sample to data) for relatively deep proteomic analysis of human biopsy samples, with a level of technical reproducibility that is comparable to S/MRM, the most highly reproducible MS-based method at present, as well as next-generation sequencing techniques. However, in contrast to S/MRM, which is limited to the quantification of maximally a few hundred proteins per sample injection, thousands of proteins can be quantified by SWATH-MS with a single injection. Moreover, it is worth noting that the technical variation we measured here for PCT-SWATH addresses the total variation from both sample preparation and MS analysis.

The unprecedented speed and precision of this proteomic analysis are achieved by seamless integration of the PCT technique with different steps of protein sample preparation and the conjugation with SWATH-MS. The method minimizes sample handling and chromatography, making it easily transferable to other laboratories given established standard operating procedures. This method also contributes to minimizing technical variation and achieving a high level of reproducibility of the MS data, with a median CV for technical replicates below 20%. Further, this method minimizes sample loss, which maximizes sample yields, even from biopsy-level tissue samples. Moreover, the simple and fast protocol makes it possible to process a large number of samples with unprecedented reproducibility, an essential factor for clinical research as it provides timely data for clinical decision making.

The resultant SWATH-MS maps contain highly reproducible signature fragment ion series for all MS-measurable peptides of a proteome within a user-defined retention time versus mass-to-charge ratio space. Unlike DDA-MS data, which are typically searched in an unbiased manner against a protein sequence database, the analysis of SWATH-MS data depends on a priori knowledge of peptide fragmentation and chromatographic behavior of the targeted peptides. Therefore, it is not possible to determine the exact number of proteins and peptides that are present in a SWATH map because noise, interfering signals and absence of some peptides in the SWATH assay library place the number of identified peptides below the maximal number of identifiable peptides. This issue cannot be fully investigated until a complete SWATH assay library for the respective sample type has been generated. However, the PCT-SWATH technology generates a comprehensive and permanent digital record of a sample from which even low-signal-intensity peptides can be reproducibly quantified and where the absence of a peptide can be verified by the absence of a signal above noise level.

Typically, we processed 1 mg of wet tissue mass. The SWATH-MS maps generated accordingly offer unique opportunities for tissue biopsy research because maps generated from the same individual over time, from different tissue samples from the same person or from tissue samples from patient cohorts can be perpetually analyzed in silico without further wet lab experimentation. This feature is particularly unique and advantageous when considering that most biopsy samples and early-stage tumor tissues are only sufficient for a limited number of analyses. Moreover, the systematic conversion of tissue samples and early-stage tumor tissues are only sufficient for a limited number of analyses. Over time, the systematic conversion of tissue samples (such as those contained in biobanks) into SWATH-MS maps could complement traditional biobanking efforts and make precious clinical samples accessible for a wide range of evaluations.

At present, there is no software tool that can deconvolute all signals in a SWATH-MS map. We therefore used the previously described targeted data analysis strategy to compare reference fragment ion spectra of proteotypic peptides generated as prior information with signal groups contained in the SWATH-MS maps. The library used here was generated by DDA analysis of 33 samples in a TripleTOF 5600 mass spectrometer and contained 49,959 reference spectra for 41,542 peptides from 4,624 reviewed Swiss-Prot proteins. This assay library applied to the OpenSWATH software allowed the quantification of 2,375 proteins across the nine paired kidney biopsies which classified the tissues into their correct histological and malignancy groups. The set of quantified proteins also contained common protein markers used in clinical and research laboratories (Fig. 4), suggesting that SWATH-MS maps, once acquired, could be used for clinical assessment. The number of proteins identified in this study was lower than the number of proteins that can be identified in studies using the most advanced DDA methods with extensive fractionation of the sample. This is in part due to the fact that in this study we focused on proteotypic peptides only, used the low redundancy Swiss-Prot database as the source of protein identities, and used a SWATH assay library of limited complexity and depth. It can be expected that from the same SWATH-MS maps, additional proteins and protein variants will be identified and quantified in the future.

The present PCT-SWATH protocol relies on in-solution digestion. Therefore, the tissue being processed should not contain contaminants that interfere with enzymatic digestion. It is common that fresh biopsy tissues are embedded in Tissue-Tek O.C.T. compound, a formulation of water-soluble glycols and resins, as the specimen matrix for cryostat sectioning. The O.C.T. must be washed away before PCT-SWATH analysis.

PCT-SWATH is a generic method for proteomic analysis that can be extended beyond biopsy tissues. PCT can be used to prepare MS-ready peptides from various types of samples, including bacteria, yeast and mammalian tissues. SWATH-MS can be used to digitize the resultant proteome. With this method, comparative quantitative analysis of a large number of proteomes is becoming possible and feasible. In the present form, PCT-SWATH is not ready to be used as a clinical test, but we expect wide applications of PCT-SWATH in personalized medicine and translational studies in general, as large-scale quantitative information on proteins is highly desirable for systems modeling and biomarker discovery. Also, if applied with the addition of carefully selected and calibrated isotope-labeled reference peptides, PCT-SWATH may reach the quantitative accuracy and reproducibility required for clinical adoption.

**METHODS**

Methods and any associated references are available in the online version of the paper.

**Accession codes.** ProteomExchange Consortium: Coordinates have been deposited with accession code PXD000672.

**Note:** Any Supplementary Information and Source Data files are available in the online version of the paper.

**ACKNOWLEDGMENTS**

We thank N. Roesch and M. Stoffel for providing test mouse tissues, H.P. Schmid and D. Engeler for their help with human tissue collection, A. Leitner and K. Novy for help in Orbitrap analysis, A. Leitner for assistance in setting up the Barocycler, the ETH Bratus team for computational support, and O.L. Kon for critical reading of the manuscript. The work was supported by the SystemsX.ch project PhosphoNetX (to R.A.), the Swiss National Science Foundation (grant no. 3100A0-688 107679 to R.A.), and the European Research Council (grant no. ERC-2008-AdG 233226 to R.A.). P.K. was supported by a...
fellowship from the Finnish Cultural Foundation. We thank the PRIDE team for support in mass spectrometry data deposition.

AUTHOR CONTRIBUTIONS
R.A. conceived the idea. T.G. developed the method. S.G., M.J. and W.J. designed the clinical cohort and collected the clinical tissue samples. T.G., P.K. and C.C.K. performed the analysis of the tissues. L.C.G. and T.G. performed the MS measurements. T.G. performed the data analysis, with critical inputs from W.E.W., C.C.K., H.L.R., G.R., B.C.C. and L.C.B. T.G. and R.A. wrote the manuscript. C.C.K., M.J. and all the other authors contributed to the revision of the manuscript. R.A. supervised the project.

COMPETING FINANCIAL INTERESTS
The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Liotta, L. & Petricoin, E. Molecular profiling of human cancer. Nat. Rev. Genet. 1, 48–56 (2000).
2. van ‘t Veer, L.J. et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 415, 530–536 (2002).
3. Sobrinho, C. & Piuzaful, L. Gene-expression signatures in breast cancer. N. Engl. J. Med. 360, 790–800 (2009).
4. Barrett, T. et al. NCBi GEO: archive for high-throughput functional genomic data. Nucleic Acids Res. 37, D885–D890 (2009).
5. Forbes, S.A. et al. COSMIC: exploring the world’s knowledge of somatic mutations in human cancer. Nucleic Acids Res. 43, D805–D811 (2015).
6. Cancer Genome Atlas Research Network, Weinstein, J.N. et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat. Genet. 45, 1113–1120 (2013).
7. Lu, J. et al. MicroRNA expression profiles classify human cancers. Nature 435, 834–838 (2005).
8. Sreekumar, A. et al. Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. Nature 457, 910–914 (2009).
9. Kononen, J. et al. Tissue microarrays for high-throughput molecular profiling of tumor specimens. Nat. Med. 4, 844–847 (1998).
10. Haab, B.B. Antibody arrays in cancer research. Mol. Cell Proteomics 4, 377–383 (2005).
11. Domon, B. & Aebersold, R. Options and considerations when selecting a quantitative proteomics strategy. Nat. Biotechnol. 28, 710–721 (2010).
12. Gillet, L.C. et al. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol. Cell Proteomics 11, 0111.016717 (2012).
13. Liu, Y. et al. Quantitative measurements of N-linked glycoproteins in human plasma by SWATH-MS. Proteomics 13, 1247–1256 (2013).
14. Collins, B.C. et al. Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14–3–3 system. Nat. Methods 10, 1246–1253 (2013).
15. Powell, B.S., Lazarev, A.V., Carlson, G., Ivanov, A.R. & Rozak, D.A. Pressure cycling technology in systems biology. Methods Mol. Biol. 881, 27–62 (2012).
16. López-Ferrer, D. et al. Application of pressurized solvents for ultrafast trypsin hydrolysis in proteomics: proteomics on the fly. J. Proteome Res. 7, 3276–3281 (2008).
17. Glatter, T. et al. Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. J. Proteome Res. 11, 5145–5156 (2012).
18. Röst, H. et al. OpenSWATH: Automated, targeted analysis of mass spectrometric data generated by data-independent acquisition. Nat. Biotechnol. 32, 219–223 (2014).
19. Rosenberger, G., Ludwig, C., Rost, H.L., Aebersold, R. & Malmstrom, L. aLFQ: an R-package for estimating absolute protein quantities from label-free LC-MS/MS proteomics data. Bioinformatics 30, 2511–2513 (2014).
20. Picotti, P., Bodenmiller, B., Mueller, L.N., Donom, B. & Aebersold, R. Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics. Cell 138, 795–806 (2009).
21. Algaba, F. et al. Current pathology keys of renal cell carcinoma. Eur. Urol. 60, 634–643 (2011).
22. Tan, P.H. et al. Renal tumors: diagnostic and prognostic biomarkers. Am. J. Surg. Pathol. 37, 1518–1531 (2013).
23. Kuster, B., Schirle, M., Mallick, P. & Aebersold, R. Scoring proteomes with proteotypic peptide probes. Nat. Rev. Mol. Cell Biol. 6, 577–583 (2005).
24. Beck, M. et al. The quantitative proteome of a human cell line. Mol. Syst. Biol. 7, 531 (2011).
25. Wiśniewski, J.R. et al. Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma. Mol. Syst. Biol. 8, 611 (2012).
26. Addona, T.A. et al. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. Nat. Biotechnol. 27, 633–641 (2009).
27. ‘t Hoen, P.A. et al. Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories. Nat. Biotechnol. 31, 1015–1022 (2013).
Online Methods

Cells and tissues. U2OS cells were obtained from the American Type Culture Collection, grown in DMEM medium supplemented with 10% bovine serum and washed three times using PBS. Mouse liver test tissues were collected from three-week-old C57BL/6J or 129Sv mice as described previously, and snap-frozen and stored at −80 °C. Prior to tissue lysis, histological analysis was performed using H&E-stained frozen sections. Tumor cell content was evaluated by a trained pathologist. All tissue samples were collected at the Institute of Pathology, Kantonsspital St. Gallen. The study protocol was approved by the ethics committee Ethikkommission St. Gallen. Written informed consent was obtained from each patient before tissue collection. Tissues embedded in OCT compounds must be washed before PCT-SWATH analysis. The OCT washing can be done according to the protocol prepared by M. Martinez and K. Shaddox, Vanderbilt-Ingram Cancer Center, which is available on the website of Clinical Proteomic Tumor Analysis Consortium (https://cptac-data-portal.georgetown.edu/cptac/dataPublic/list?currentPage=2%2FPhase_II_Data%2FOCT%2FOCT_Embedded_Tumor_Tissue_VU&nonav=true). Briefly, tissue is placed in a 1.5-ml microcentrifuge tube and vortexed in the following buffers in sequence: (1) 1 ml 70% ethanol/30% water for 30 s; (2) 1 ml 100% water for 30 s; (3) 1 ml 70% ethanol/30% water for 5 min, twice; (4) 1 ml 85% ethanol/15% water for 5 min, twice; (5) 1 ml 100% ethanol for 5 min, twice.

PCT-based lysis and digestion. U2OS cells, mouse liver and human kidney tissues were digested using a Barocycler NEP2320-45k (PressureBioSciences, South East, MA). Cell pellets (cell count range, 300,000–900,000) and mechanically disrupted tissue pieces (weight range: 1–3 mg) were placed in MicroTubes (PressureBioSciences) with lysis buffer consisting of 8 M urea, 0.1 M ammonium bicarbonate, oComplete protease inhibitor cocktail (Roche) and PhosSTOP phosphatase inhibitor cocktail (Roche). The solution was sonicated for 20 s on ice and placed in the Barocycler for tissue lysis and protein extraction at 33 °C. The cyling process was programmed to contain 60 cycles, each consisting of 50 s at 45,000 p.s.i. high and 10 s at ambient pressure. The protein solution was transferred to new MicroTubes for subsequent processes. Tris(2-carboxyethyl)phosphine and iodoacetamide were simultaneously added to the solution at a final concentration of 10 mM and 40 mM, respectively, followed by a 30-min incubation in the dark with gentle vortexing. The 8 M urea was diluted to 6 M using 0.1 M ammonium bicarbonate. Proteins were first digested using Lys·C (Wako; enzyme-to-substrate ratio, 1:40) in the Barocycler at 33 °C. The cycling process was programmed to contain 45 cycles, each consisting of 50 s at 20,000 p.s.i. and 10 s at ambient pressure. Tryptic digestion was then performed after diluting the urea concentration to 1.6 M using 0.1 M ammonium bicarbonate. Trypsin (Promega) was used with an enzyme-to-substrate ratio of 1:20. Tryptic digestion was performed in the Barocycler at 33 °C using 90 cycles of 50 s at 20,000 p.s.i. and 10 s at ambient pressure. Subsequently, trifluoroacetic acid was added to the solution at a final concentration of 0.4% to stop digestion. The peptides were cleaned using SEP-PAK C18 cartridges (Waters Corp., Milford, MA), dried under vacuum and stored at −80 °C before MS analysis. Peptides were dissolved in HPLC-grade water containing 0.1% formic acid and 2% acetonitrile. To facilitate peptide solubilization, peptide solution was sonicated in an icy water bath for 5 min. Peptide concentration was measured using NanoDrop 1000 spectrophotometer with the option of protein A280 (1 Ab = 1 mg ml−1). Peptide concentration was adjusted to 0.3 µg µl−1 using 0.1% formic acid and 2% acetonitrile in HPLC water. 1 µl of iRT peptide cocktail was spiked into every 9 µl of peptide sample before LC-MS analysis.

Orbitrap MS analysis. The peptide samples were analyzed on an LTQ Orbitrap XL mass spectrometer (Thermo Scientific) using Xcalibur version 2.0.7 coupled to an Eksigent NanoLC-2D HPLC nanoflow system (single-pump system with one analytical column; Eksigent) via a nanoelectrospray ion source using a liquid junction (Thermo Scientific), as described previously. The raw data were converted to mzXML format and searched using Sorcerer against a target-decoy Swiss-Prot mouse protein database. False discovery rate was controlled below 1% using Trans-Proteomic Pipeline (TPP, version 4.6.1).

Triple-TOF MS analysis in DDA and SWATH mode. A 5800 TripleTOF mass spectrometer (AB Sciex) and an Eksigent 1D+ Nano LC systems (Eksigent, Dublin, CA) were employed for SWATH-MS analysis. The LC gradient was formulated with buffer A (2% acetonitrile and 0.1% formic acid in HPLC water) and buffer B (2% water and 0.1% formic acid in acetonitrile). The analytical column (75 µm × 20 cm) was home-packed directly in a fused silica PicoTip emitter (New Objective, Woburn, MA, USA) with 3-µm 200 Å Magic C18 AQ resin (Michrom BioResources, Auburn, CA, USA). The peptide samples were prepared at 0.3 µg µl−1 spiked-in with 1:20 (vol/vol) iRT peptides. 3 µl of each sample were injected and separated with a linear gradient of 2% to 35% buffer B over 120 min at a flow rate of 0.3 µl min−1. The column was then flushed with 90% buffer B for 5 min, and reequilibrated with 2% buffer B for 10 min. For shotgun experiments, the TripleTOF mass spectrometer was operated with a ‘top 20’ method. From the 500-ms survey scan (TOF-MS), the 20 most intense precursor ions were selected for subsequent automated MS/MS in measurements wherein each MS/MS event consisted of a 150-ms scan. The selection criteria for parent ions included intensity, where ions had to be greater than 200 counts per second (c.p.s.) and have a charge state greater than 2+. The precursors that were selected for fragmentation were added to a dynamic exclusion list for 20 s. Ions were isolated using a quadrupole resolution of 0.7 AMU and fragmented in the collision cell using the collision energy equation (0.0625 × m/z − 3.5) with an additional collision energy spread of 15 eV within the 150-ms accumulation time to mimic SWATH fragmentation conditions. In the instances where there were ≤20 parent ions per survey scan that met the selection criteria, those ions were subjected to extended MS/MS accumulation times to maintain a constant total cycle time of 3.5 s. For SWATH-MS based experiments, the mass spectrometer was operated in a looped product ion mode. In this mode the instrument was specifically tuned to allow a quadrupole resolution of 25 AMU per mass selection. The stability of the mass selection was maintained by the operation of the RF and DC voltages on the isolation quadrupole in an independent manner as described earlier. Using an isolation width of 25 AMU, a set of 32 overlapping windows (1 AMU overlap) was constructed, covering the mass range of 400–1,200 AMU. The collision energy for each window was determined on the basis of the collision energy for a 2+ ion centered in the respective window (equation: 0.0625 × m/z − 3.5) with a collision energy spread of 15 eV. An accumulation time of 100 ms was used for each fragment ion scan and for the survey scans acquired at the beginning of each cycle, resulting in a total cycle time of 3.3 s. The sequential precursor isolation window setup was as follows: 400–425, 425–450, 449–475, 474–500, 499–525, 524–550, 549–575, 574–600, 599–625, 624–650, 649–675, 674–700, 699–725, 724–750, 749–775, 774–800, 799–825, 824–850, 849–875, 874–900, 899–925, 924–950, 949–975, 974–1000, 999–1025, 1024–1050, 1049–1075, 1074–1100, 1099–1125, 1124–1150, 1149–1175, 1174–1200, with an effective (100%) transmission of 25 AMU and 0.3 AMU attraction on either side of the isolation window.

Kidney tissue SWATH assay library generation. A pool of the 18 kidney tissue samples used for SWATH analysis was resolved into 15 gel fractions using SDS–PAGE and digested individually in gel. The peptides extracted from each gel band were subjected to DDA analysis in the TripleTOF MS instrument. The individual digests of the whole proteomes of the 18 kidney tissues were also analyzed in TripleTOF in DDA mode. iRT peptides, a set of artificial peptides synthesized for retention time calibration, were spiked to these samples at a concentration of 10% before LC-MS analysis. All 33 DDA data files were searched for peptide identification using X!Tandem and OMSSA against a target-decoy, nonredundant human UniProtKB/Swiss-Prot protein database (Oct 1, 2013) containing 20,265 protein sequences and the iRT peptide sequences. The decoy sequences were generated by reversing each of the target sequences. Fully tryptic peptides with maximal two missed cleavages were allowed. The mass tolerance of parent mass was set to 50 p.p.m. Fragment mass tolerance was set at 0.1 Da. Carbamidomethyl at cysteine was set as static modification, and oxidation at C terminal Met was set as variable modification. }
methionine was set as variable modification. Search results from X!Tandem and OMSA were further analyzed through Trans-Proteomic Pipeline (TPP, version 4.6.0)\(^{19}\) using PeptideProphet and iProphet. The cutoff of probability for peptides was 0.9. The resultant pep.xml files were combined into a single file named as interact.pep.xml.

SpectraSt (version 4.0) in TPP and a previously described workflow\(^{18}\) were used to process the resulting interact.pep.xml file, normalize the retention time, generate consensus spectra and produce the TraML library file following the commands below:

```
spectrast -o step2.splib -c BIN! -cf 'Protein! - DECOY - cP0.9 - cICID-QTOF - cNstep1 interact.pep.xml
spectrast2spectrast_irt.py -i step1.splib -o step2.splib
spectrast -cAC -c BIN! -cf 'Protein! - iRT' - cNstep3 step2.splib
spectrast2tsv.py -l 400,2000 -s b,y -x 1,2,3,4 -o 6 -n 6 -p 0.05 -e -w swaths.txt -k openswath -a step4.tsv step3.splib
mv step4.tsv step4.csv
```

All the tools used here are available under the modified BSD license at http://www. swath对外开放.org/ and https://code.google.com/p/msproteomicstools/. Documentation for OpenSWATH and a detailed protocol for SWATH assay library generation is available\(^{18}\). In our laboratory, the entire pipeline was streamlined as an iPortal workflow\(^{34}\).

**SWATH data analysis.** SWATH wiff files were converted into mzXML files using ProteoWizard msconvert\(^{35}\) v.3.0.3316. A Python script (fix_swath_windows.py) was used to correct precursor isolation-window mass error introduced by mconvert as described previously. The mzXML file was converted to mzML file using OpenMS tool FileConverter\(^{36}\). OpenSWATH was performed using the tool OpenSWATHWorkflow with input files including the mzXML file, the TraML library file and the TraML file for iRT peptides\(^{18}\). The assays were identified with a false discovery rate below 0.1%. High confidence peptide features from different samples were aligned using the algorithm TRansition Identification Confidence (TRIC) (version v238), which is available from https://pypro.python.org/ippit/msproteomicstools or https://code.google.com/ p/msproteomicstools/. The following parameters for the feature_alignment.py are as follows: max Rt_diff = 30, method = global_best_overall, nr_high_conf_exp = 2, target_fdr = 0.001, use_score_filter = 1.

Label-free protein-level quantification was performed using the function ProteinInference from the R package alFQ (version 1.3.1)\(^{19}\) on output from OpenSWATH. The parameters that differ from the default for ProteinInference are: peptide_method = "top", peptide_topx = 3, transition_topx = 5, consensus_proteins = FALSE, consensus_peptides = FALSE, consensus_transitions = FALSE.

The MS proteomics data and SWATH assay library have been deposited to the ProteomeXchange Consortium (http://www.proteomexchange.org/) via the PRIDE partner repository\(^{37}\) with the data set identifier PXD000672.

**Statistical analysis.** The protein expression values were subjected to the arcsin variance stabilizing transformation\(^{38}\), and the effect was assessed by Bland–Altman diagnostic plotting. We corrected for systematic differences in expression levels between the samples using the quantile normalization\(^{39}\). Missing values were excluded from calculation of CV values, but considered as 0 in heat map generation and t-tests. Variation analysis of the RCC SWATH data was performed by comparing quantified proteins between each pair of SWATH data sets, taking into account all combinations. The median CV of all proteins was plotted using the R package vioplot. The MS injection variation analysis considered 18 pairs of data sets. The within-patient variation analysis of all RCC and ccRCC considered 36 and 24 pairs of data sets, respectively. Regarding inter-patient variation analysis in all RCC patients, one data set was sampled from each of the 9 patients and pooled for variation analysis. For all RCC tissues, altogether 262,144 combinations were analyzed. 512 combinations were included for nontumorous and tumorous tissues from all RCC patients. Variation analysis of ccRCC was based on 4,096 combinations of ccRCC data sets, and 64 combinations of nontumorous or tumorous ccRCC data sets. Heat maps and violin plots were plotted using R. P values in the volcano plots were calculated using paired or unpaired two-tailed t-tests as indicated in the respective legend.

28. Sabido, E. et al. Targeted proteomics reveals strain-specific changes in the mouse insulin and central metabolic pathways after a sustained high-fat diet. Mol. Syst. Biol. 16, 681 (2013).
29. Escher, C. et al. Using iRT, a normalized retention time for more targeted measurement of peptides. Proteomics 12, 1111–1121 (2012).
30. Leitner, A. et al. Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. Mol. Cell. Proteomics 11, M111.014126 (2012).
31. MacLean, B., Eng, J.K., Beavis, R.C. & McIntosh, M. General framework for developing and evaluating database scoring algorithms using the TANDEM search engine. Bioinformatics 22, 2830–2832 (2006).
32. Geer, E.Y. et al. Open mass spectrometry search algorithm. J. Proteome Res. 3, 958–964 (2004).
33. Deutsch, E.W. et al. A guided tour of the Trans-Proteomic Pipeline. Proteomics 10, 1150–1159 (2010).
34. Kessner, D., Chambers, M., Burke, R., Agus, D. & Mallick, P. ProteoWizard: iPortal: the Swiss Grid Proteomics Portal: requirements and new features based on experience and usability considerations. Concurreny and Computation: Practice and Experience 27, 433–445 (2015).
35. Kunszt, P. et al. IPortal: the Swiss Grid Proteomics Portal: requirements and new features based on experience and usability considerations. Concurreny and Computation: Practice and Experience 27, 433–445 (2015).
36. Leitner, A. et al. OpenMS: an open-source software framework for mass spectrometry. BMC Bioinformatics 9, 163 (2008).
37. Vizcaíno, J.A. et al. The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res. 41, D1063–D1069 (2013).
38. Huber, W., von Heydebreck, A., Sultmann, H., Poustka, A. & Vingron, M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics 18 (suppl. 1), S96–S104 (2002).
39. Bolstad, B.M., Irizarry, R.A., Astrand, M. & Speed, T.P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19, 185–193 (2003).