

# A quantitative and site-specific atlas of the citrullinome reveals widespread existence of citrullination and insights into PADI4 substrates

Received: 22 December 2022

Accepted: 4 January 2024

Published online: 6 February 2024

 Check for updates

Alexandra S. Rebak<sup>1</sup>, Ivo A. Hendriks<sup>1</sup>, Jonas D. Elsborg<sup>1</sup>, Sara C. Buch-Larsen<sup>1</sup>, Claus H. Nielsen<sup>2</sup>, Lene Terslev<sup>3</sup>, Rebecca Kirsch<sup>4</sup>, Dres Damgaard<sup>2</sup>, Nadezhda T. Doncheva<sup>4</sup>, Caroline Lennartsson<sup>1</sup>, Martin Rykær<sup>1</sup>, Lars J. Jensen<sup>4</sup>, Maria A. Christophorou<sup>5</sup> & Michael L. Nielsen<sup>1</sup>✉

Despite the importance of citrullination in physiology and disease, global identification of citrullinated proteins, and the precise targeted sites, has remained challenging. Here we employed quantitative-mass-spectrometry-based proteomics to generate a comprehensive atlas of citrullination sites within the HL60 leukemia cell line following differentiation into neutrophil-like cells. We identified 14,056 citrullination sites within 4,008 proteins and quantified their regulation upon inhibition of the citrullinating enzyme PADI4. With this resource, we provide quantitative and site-specific information on thousands of PADI4 substrates, including signature histone marks and transcriptional regulators. Additionally, using peptide microarrays, we demonstrate the potential clinical relevance of certain identified sites, through distinct reactivities of antibodies contained in synovial fluid from anti-CCP-positive and anti-CCP-negative people with rheumatoid arthritis. Collectively, we describe the human citrullinome at a systems-wide level, provide a resource for understanding citrullination at the mechanistic level and link the identified targeted sites to rheumatoid arthritis.

Post-translational modifications (PTMs) are chemical changes that occur on proteins in response to cell stimuli. They are often reversible and act as dynamic molecular switches that modulate protein structure and function, thereby adding enormous complexity to the proteome, and enhance the regulatory potential of cells.

Arginine is converted to citrulline through citrullination, a process that is facilitated by enzymes known as peptidylarginine deiminases (PADIs or PADs). This modification, considered irreversible, involves hydrolyzing the guanidinium group of the arginine side chain, affecting the charge and hydrogen-bonding ability of the amino acid.

<sup>1</sup>Proteomics Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. <sup>2</sup>Institute for Inflammation Research, Center for Rheumatology and Spine Diseases, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark. <sup>3</sup>Copenhagen Center for Arthritis Research, Center for Rheumatology and Spine Diseases, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark. <sup>4</sup>Disease Systems Biology Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. <sup>5</sup>Epigenetics, Babraham Institute, Cambridge, UK. ✉e-mail: [Michael.Lund.Nielsen@cpr.ku.dk](mailto:Michael.Lund.Nielsen@cpr.ku.dk)

Humans have five PADI isoforms (PADI1–PADI6), each of which has different tissue and cellular distributions<sup>1,2</sup>. Citrullination can affect interactions with nucleic acids and proteins, potentially altering protein structures<sup>3–5</sup>.

Citrullination has functional roles in various biological processes<sup>5–9</sup>. Evidence suggests that it is involved in various disorders, including autoimmunity, neurodegeneration, atherosclerosis and cancer<sup>10–12</sup>.

Citrullination in neutrophils is attracting increasing attention as a mediator of rheumatoid arthritis (RA)<sup>13,14</sup> and systemic lupus erythematosus (SLE)<sup>15,16</sup>. In RA, increased protein citrullination in inflamed joints, the presence of autoantibodies (ACPAs) and the genetic association of PADI4 with RA development underscore the importance of this process<sup>17–19</sup>. ACPAs, detected through anti-cyclic citrullinate peptide/protein (CCP) assays, are diagnostic and prognostic markers of RA<sup>20</sup>. Inhibition of PADI activity has shown promise in disease models, making it a potential therapeutic strategy for RA, ulcerative colitis, neurodegeneration and cancer<sup>2,21</sup>.

Despite great biological and clinical interest, the understanding of citrullination and its associated molecular mechanisms remains limited. Mapping citrullinated proteins and modification sites has been challenging and limited by contemporary analytical strategies. Robustly sensitive enrichment methods are lacking, and mass spectrometers are unable to discern specific, low-abundance citrullination events in complex cell extracts<sup>22–27</sup>.

To comprehensively map citrullination with high precision, we used advanced proteomics technology. Our approach involved offline high-pH reversed-phase fractionation<sup>28</sup> followed by online low-pH reversed-phase fractionation, coupled with high-resolution mass spectrometry (MS). We delineated the human citrullinome and identified specific citrullination sites catalyzed by PADI4. To capture the functional and biological aspects of the process, we used HL60 promyeloblasts as a model for human neutrophil behavior. Following DMSO treatment, these cells differentiated into neutrophil-like cells (NLCs), increasing PADI4 enzyme expression<sup>29</sup>. We induced PADI enzyme activity by elevating cellular calcium ion concentrations and assessed citrullination catalysis through quantitative MS.

Collectively, our analyses identified 14,056 high-confidence citrullination sites on 4,008 proteins, vastly expanding the number of known sites by >16-fold. Identified citrullination events occurred on core histone variants and non-histone proteins, broadening our understanding of citrullination's regulatory functions and underscoring the current underestimation of citrullinated autoantigens in neutrophils. Using a peptide microarray, we investigated ACPA specificity in synovial fluid from people with RA or ankylosing spondylitis, revealing differences in reactivity and binding motifs. Collectively, our findings demonstrate the extensive regulatory role of citrullination, showcasing how MS data can enhance our understanding of autoantigens.

## Results

To identify citrullinated proteins and their modification sites globally, we used a proteomics-based approach with human promyelocytic leukemia cells (HL60) with or without calcium ionophore. Recognizing the challenge of detecting low-abundance citrullination sites in complex samples<sup>30</sup>, we implemented offline high-pH reversed-phase fractionation into 46 fractions to reduce sample complexity<sup>28</sup> (Fig. 1a). This strategy, combined with the high sequencing speed of modern MS instrumentation<sup>31</sup>, enabled direct analysis of citrullination events, without needing PTM-specific enrichment.

As a first step, we confirmed by western blot that DMSO had induced differentiation of HL60 cells into NLCs. The blot revealed that the PADI4 enzyme was upregulated (Fig. 1b). The addition of a calcium ionophore enhances the catalytic activity of PADI4. Because prolonged treatment can recapitulate neutrophilic extracellular trap (NET) formation *in vitro*<sup>32</sup>, we opted for 15- and 30-min treatments, which resulted in a total of five biological conditions: (1) HL60, no calcium; (2) HL60, 15 min calcium; (3) NLCs, no calcium; (4) NLCs, 15 min calcium; and (5) NLCs, 30 min calcium. The western blot detected profound global citrullination, including citrullinated histone H3R2 (H3R2Cit)<sup>33</sup>, a known citrullination target, as well as modified citrulline, detected by a pan-peptidyl-citrulline antibody (Mod Cit) (Fig. 1b). For improved identification and localization of citrullination sites, samples were digested with endoproteinase Lys-C, ensuring that the arginine residues targeted for citrullination were positioned internally in the analyzed peptide sequences. Each sample fraction was analyzed on short gradients using high-resolution MS (Q-Exactive HF-X)<sup>34</sup>. This allowed the citrullinome to be measured in less than 3 d. Quicker analysis could be achieved by analyzing fewer fractions. In total, we analyzed 138 liquid chromatography–tandem MS (LC–MS/MS) runs for each condition, and performed rigorous and stringent data filtering using MaxQuant to ensure that identification and quantification of citrullinated peptides was robust, with a high degree of confidence.

### Proteome changes upon PADI activation

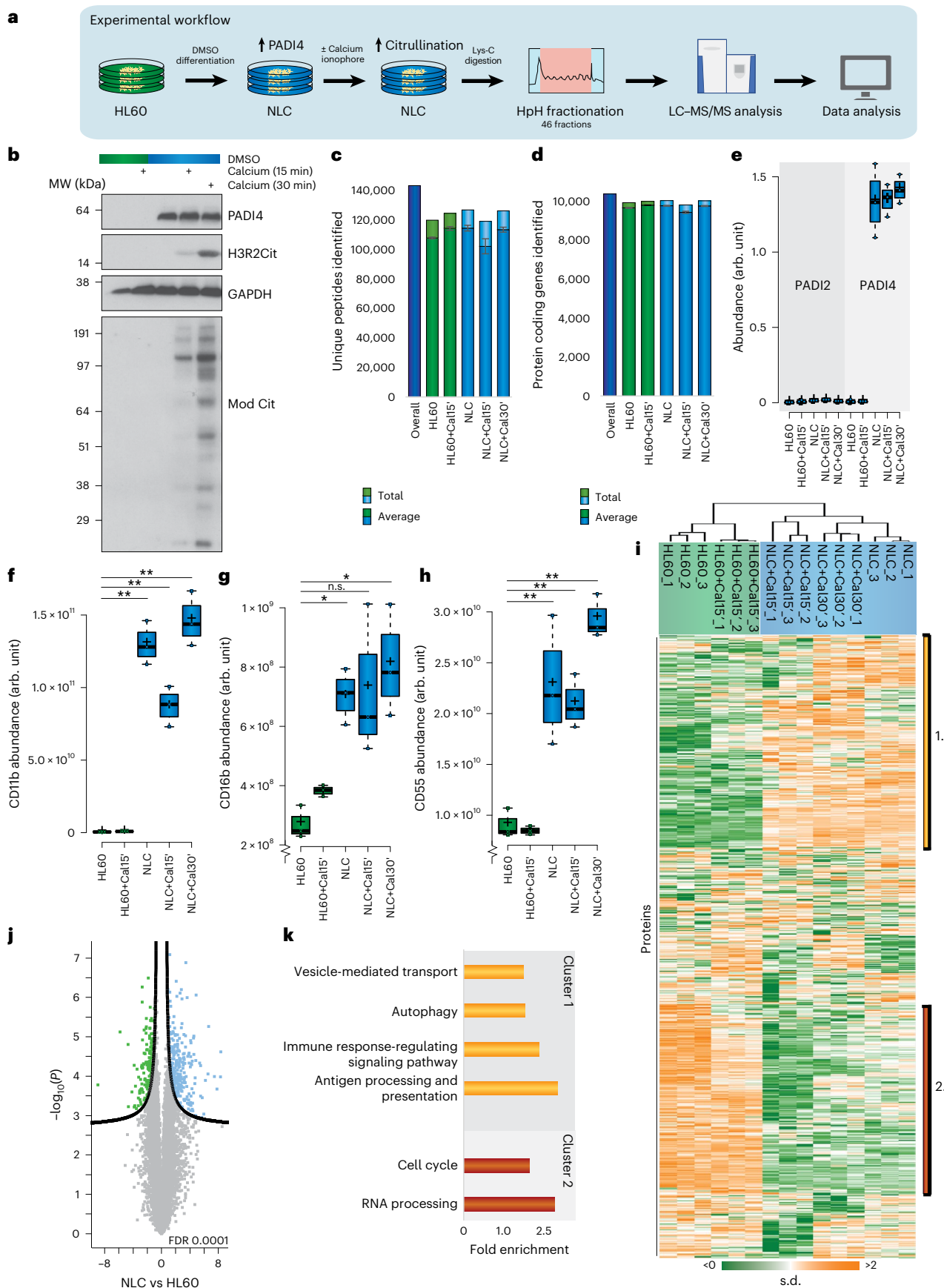
Because we performed high-pH fractionation on whole-cell lysate and did not perform a PTM-specific enrichment step<sup>28</sup>, our proteomics strategy allows for investigation of cellular protein-expression changes (that is, the proteome) concomitantly with the profiling of the citrullinome<sup>28</sup>. Overall, the biological replicates of the proteome measurements demonstrated high quantification accuracy and reproducibility, with a clear distinction between HL60 and NLC cells (Extended Data Fig. 1a) with Pearson correlation coefficients of 0.97–0.99 between replicate analyses (Extended Data Fig. 1b). We identified a total of 143,403 unique peptides, with a false discovery rate (FDR) of 1% across experimental conditions; in each individual condition, >100,000 unique peptides were identified (Fig. 1c and Supplementary Table 1). The identified peptides mapped onto a total of 10,630 unique protein-coding genes,

**Fig. 1 | Establishment of regulatory citrullination cell system.** **a**, Schematic of the workflow. HL60 cells were cultured in standard RPMI medium or differentiated by DMSO into neutrophil-like cells (NLCs). Cells were then incubated in Locke's solution with 4  $\mu$ M calcium ionophore for 0, 15 or 30 min. Following Lys-C digestion, the peptides were separated by high-pH (HpH) fractionation and analyzed using high-resolution LC–MS/MS. Each condition was prepared in triplicate. **b**, Western blot analysis of PADI4, H3R2Cit and modified citrulline shows that PADI4 is expressed upon DMSO differentiation and activated following calcium treatment. The experiment was repeated three times independently, with similar results. The modified citrulline was run on a different gel than was the GAPDH loading control, owing to different sample-preparation requirements. MW, molecular weight. **c**, The number of cumulative peptides detected and across the conditions.  $n = 3$  biologically independent samples. Data are presented as mean values  $\pm$  s.d. Cal15' and Cal30', calcium treatment for 15 or 30 min, respectively. **d**, The number of cumulative proteins detected and across the conditions.  $n = 3$  biologically independent samples. Data are presented as

mean values  $\pm$  s.d. **e**, Abundance of PADI2 and PADI4 across conditions, based on label-free quantification (LFQ) intensity.  $n = 3$  biologically independent samples; whiskers, 95th and 5th percentiles; box limits, third and first quartiles; center bar, median; '+' symbol, average. arb. unit, arbitrary unit. **f**, As in **e**, but for CD11b. **g**, As in **e**, but for CD16b. **h**, As in **e**, but for CD55. **i**, Hierarchical clustering analysis of z-scored intensity values of all proteins detected across all conditions. Orange, upregulated; green, downregulated. Bars 1 and 2, correspond to clusters 1 and 2, respectively, as analyzed in **k**. **j**, Volcano plot for visualization of protein dynamics in response to DMSO differentiation, from the HL60 to NLC condition. FDR, 0.0001. **k**, GO term enrichment analysis for biological processes of citrullination target proteins of selections displayed in **d**, as compared with the detected HL60 proteome. Statistical information is available in Supplementary Table 5. For **f–h**, significance was determined by two-tailed Student's *t* test with no correction for multiple comparisons. Asterisks indicate significant difference between HL60 and the indicated conditions: \* $P < 0.05$ ; \*\* $P < 0.001$ ; n.s., non-significant.

and more than 9,400 unique protein-coding genes were quantified across individual experiments (Fig. 1d and Supplementary Table 2). The capacity of high-pH fractionation to improve peptide sequence

coverage of identified proteins has been demonstrated<sup>28</sup>. Likewise, our analysis yields a median sequence coverage of 37% for the identified proteins (Extended Data Fig. 1c).



Consistent with our observations using western blotting (Fig. 1b) and literature reports<sup>29</sup>, our proteome analysis confirms that expression of PADI4 is significantly induced upon the differentiation of HL60 cells into NLCs. These observations align with high expression levels of PADI4 in neutrophils<sup>35</sup>. Notably, the increased expression of PADI4 occurred before cells were treated with calcium ionophore, confirming that PADI4 upregulation was not merely an indirect consequence of ionophore stimulation (Fig. 1e). Neutrophils are known to express PADI4 and PADI2 (refs. 35,36). We found evidence for low expression of PADI2, but no changes in expression were observed across the investigated cellular conditions (Fig. 1e). Despite our deep proteome analysis, we found no evidence for expression of PADI enzymes besides PADI4 and PADI2 (Supplementary Table 2). We conclude that PADI2 and PADI4 are likely the enzymes that are responsible for citrullination catalysis in the analyzed cells. Expression levels of previously reported neutrophil markers, including CD11b (Fig. 1f), CD16b (Fig. 1g) and CD55 (Fig. 1h)<sup>37,38</sup>, were increased in expression upon differentiation into NLCs.

Our protein-expression data enable a systems-level assessment of the effects of DMSO treatment on the proteome. Our data set provides a foundation for understanding the global protein-expression patterns underlying functional specialization of HL60 cells into NLCs. To obtain a functional classification of the differences related to differentiation, we performed unsupervised hierarchical clustering of the >10,000 identified proteins (Fig. 1i). The resulting heat map revealed one major cluster of proteins that is highly expressed in differentiated NLCs (Fig. 1i,j), with Gene Ontology (GO) analysis revealing enrichment ( $P < 0.006$ ) of terms related to neutrophil functions, including immune-response-regulating pathways, autophagy, antigen processing and presentation, and vesicle-mediated transport (Fig. 1k and Extended Data Fig. 1d). Contrary to this, a second major cluster (cluster 2) comprised proteins with high expression in undifferentiated HL60 cells; it was enriched in biological processes such as cell cycle and RNA processing. In summary, our results affirm the differentiation of HL60 cells into NLCs, establishing relevant biological conditions conducive for studying citrullination events mediated by the PADI4 enzyme.

### Proteome-wide identification of citrullination sites

Having established that NLCs express PADI4, we next evaluated the capability of our methodology to identify citrullination sites. Overall, we observed a high degree of reproducibility across biological replicates of NLCs (Extended Data Fig. 2a), which resulted in ~70% overlap in citrullination sites between any two replicate runs of NLCs (Extended Data Fig. 2b). For all investigated cells and treatments, a principal component analysis (PCA) analysis revealed clear separation between the individual cell populations (Fig. 2a). Reassuringly, the high precursor-mass accuracy of the MS analysis (<2 mDa) ensured that the mass increment caused by citrullination (-0.9840 Da) could be easily distinguished from the change in peptide mass resulting from the naturally occurring stable isotopes of carbon and nitrogen (<sup>13</sup>C, +1.0034 Da; <sup>15</sup>N, +0.9970 Da)<sup>39</sup> (Fig. 2b,c).

To determine the location of citrullination sites within the identified peptides, we used high-resolution MS/MS analysis. Each identified spectrum was assigned a localization score to indicate the confidence

level associated with the identification of the amino acid harboring the citrullination modification (Fig. 2d and Extended Data Fig. 2d). Because citrullination of arginine residues results in the same peptide mass shift as deamidation of glutamine and asparagine residues, we considered only citrullination sites identified from MS/MS spectra with high localization probability (>0.9) (Extended Data Fig. 2c), thereby ensuring high confidence in distinguishing between amino acids that become citrullinated or deamidated. The average localization probability of all identified citrullination sites in our filtered data set was 99.41%.

To further confirm our ability to differentiate citrullination from deamidation, we investigated the chromatographic elution times of citrullinated peptides and compared these with those of unmodified peptides and deamidated peptides. We found that there was distinct chromatographic separation between citrullinated peptides and their unmodified counterparts (Fig. 2e and Extended Data Fig. 2e), that citrullinated peptides and deamidated peptides had different *m/z* ratios (Extended Data Fig. 2f) and that there was no spatial bias of citrullination site location relative to asparagine (N) and glutamine (Q) targets for deamidation (Extended Data Fig. 2g), confirming robust differentiation between citrullination and deamidation events (see Supplementary Note 1 for further details).

Our pilot analysis resulted in high-confidence identification of 5,238 citrullination sites across the investigated cell conditions (Supplementary Table 3). Within these data, we investigated which proteins were most abundantly modified with citrullination, regardless of the total number of modification sites, encompassing proteins associated with various biological processes (Table 1). The highest number of citrullination sites was identified in NLCs after 30-min stimulation with calcium ionophore (Fig. 2g), in agreement with western blot observations (Fig. 1b). Reassuringly, we observed that >75% of the acquired citrullinated MS/MS spectra (Extended Data Fig. 2i) exhibited a consistent neutral loss of isocyanic acid, representing ~95% of the total citrullinated peptide abundance (Extended Data Fig. 2j). This distinctive characteristic serves as a reliable signature for the confident identification of citrullination sites<sup>40</sup>. In contrast to the increased citrullination abundance in NLCs (Fig. 2f), we observed no significant change in the number of deamidation sites (Extended Data Fig. 2h) across experimental conditions. This observation underscores that our high-resolution proteomics data allows citrullination to be distinguished from deamidation. We found that the increase in citrullination-site abundance between 15 and 30 min of calcium ionophore treatment (Fig. 2f) was higher than the corresponding increase in the number of citrullination sites for the same conditions (Fig. 2g). This indicates that PADI4 activation is specific and tightly regulated, with longer periods of calcium ionophore stimulation leading to enhanced citrullination of the same arginine residues.

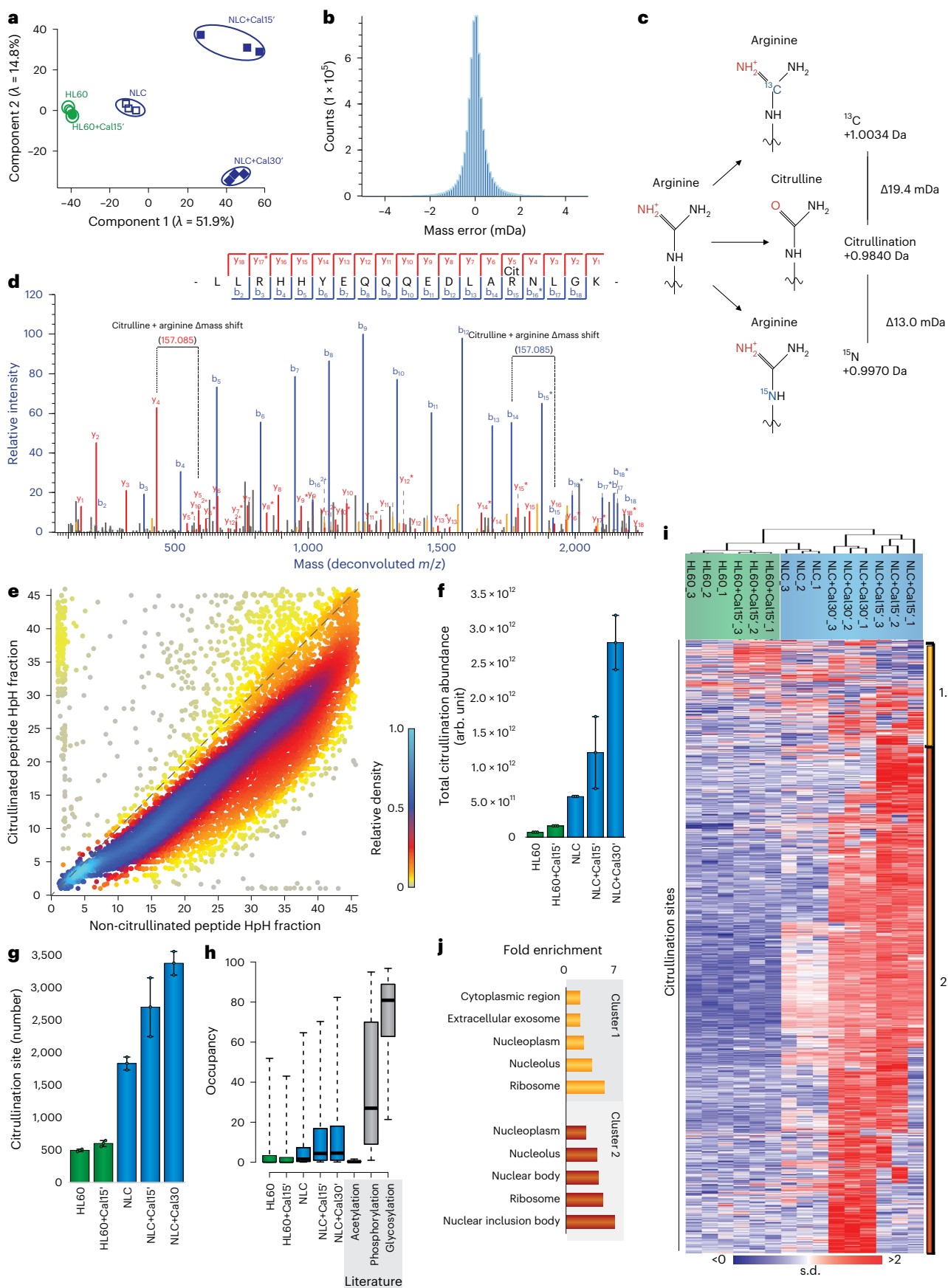
Next, we wanted to assess the 'citrullination modification stoichiometry' (that is, the modification percentage of any given arginine residue, also referred to as occupancy), because such information is valuable for understanding the implications of PTMs in protein regulation. Although stoichiometry in itself does not provide direct evidence of biologically relevant function, it is expected that sites with higher modification stoichiometries are more likely to have functional

**Fig. 2 | Validation of regulatory citrullination cell system.** **a**, PCA of all conditions at the citrullination-site level. Eigenvalues are displayed on the axes. **b**, Mass accuracy of all identified peptides. **c**, Diagram displaying small mass shifts between citrullination, carbon-13 (<sup>13</sup>C) and nitrogen-15 (<sup>15</sup>N) substitution on residue. **d**, Representative MS/MS spectrum, showing localized citrullination. **e**, Density plot of the fraction distribution during high-pH reversed-phase separation for unmodified peptides (*x* axis) compared with citrullinated peptides (*y* axis). **f**, Total citrullination abundance of sites detected across conditions. *n* = 3 biologically independent samples. Data are presented as mean values ± s.d. **g**, Number of citrullination sites detected in the five conditions by MS/MS and

matching. Localization score > 0.75. *n* = 3 biological independent samples. Data are presented as mean values ± s.d. **h**, Occupancy of citrullination across the five conditions. For comparison, occupancies of acetylation<sup>90</sup>, phosphorylation<sup>91</sup> and glycosylation<sup>19</sup> are shown. Whiskers, 95th and 5th percentiles; box limits, third and first quartiles; center bar, median; '+' symbol, average. **i**, Hierarchical clustering analysis of *z*-scored intensity values of all proteins detected across all conditions. Red, upregulated; blue, downregulated. **j**, Gene Ontology term enrichment analysis for biological process of citrullination target proteins in HL60 cells and the NLC+Cal30' condition, in addition to selections displayed in **i** (clusters 1 and 2), as compared with the human proteome. Statistical information is available in Supplementary Table 5.

consequences<sup>41</sup>. Across the citrullinome, we observed that the modification stoichiometry increased from -7% for the HL60 cells to -16% for the NLC cells upon DMSO-induced differentiation and 30-min calcium

ionophore treatment (Fig. 2h). We compared citrullination stoichiometry with the previously published stoichiometry of other PTMs, and found that it was larger than the stoichiometry of acetylation yet lower



**Table 1 | Top 50 most abundantly citrullinated proteins identified in system validation study on the HF-Exactive system**

Gene name	Total number of sites	Most abundant site	Second most abundant site	Third most abundant site	Biological process
<i>H3C1</i>	1	R27	–	–	Chromatin organization
<i>ALYREF</i>	3	R141	R144	R253	mRNA processing
<i>MNDA</i>	3	R127	R119	R129	Activation of innate immune response
<i>RPL24</i>	9	R92	R140	R88	mRNA processing
<i>H3-3A</i>	6	R9	R27	R18	Telomere organization
<i>RSL24D1</i>	10	R157	R80	R19	Translational initiation
<i>NIPSNAP3B</i>	2	R96	R100	–	Mitophagy regulation
<i>ENO1</i>	6	R400	R403	R50	Canonical glycolysis
<i>H1-4</i>	2	R54	R33	–	Chromatin organization
<i>H2AC4</i>	3	R43	R4	R72	Chromatin organization
<i>HNRNPA3</i>	9	R113	R377	R376	mRNA processing
<i>H1-10</i>	6	R128	R127	R155	Chromatin organization
<i>SUB1</i>	2	R47	R59	–	Negative regulation of DNA metabolic process
<i>HMGB1</i>	4	R24	R73	R70	Activation of innate immune response
<i>HNRNPA2B1</i>	11	R352	R21	R350	mRNA processing
<i>NPM1</i>	6	R197	R142	R277	Cell aging
<i>RPL29</i>	7	R95	R18	R68	Translational initiation
<i>SERF2</i>	7	R26	R7	R3	Protein destabilization
<i>CORO1A</i>	7	R3	R12	R402	Actin filament organization
<i>TAGLN2</i>	6	R12	R4	R160	Platelet degranulation
<i>HNRNPM</i>	5	R56	R60	R707	mRNA processing
<i>PTMA</i>	1	R90	–	–	Transcription
<i>RPL10</i>	8	R24	R203	R21	Translational initiation
<i>RPS10</i>	7	R124	R153	R119	Translational initiation
<i>DEK</i>	10	R75	R8	R65	Chromatin organization
<i>MSN</i>	6	R503	R408	R393	Cellular response to testosterone stimulus
<i>SRSF3</i>	1	R3	–	–	Cellular response to leukemia inhibitory factor
<i>LBR</i>	9	R141	R96	R95	Neutrophil differentiation
<i>HNRNPU</i>	12	R255	R268	R250	mRNA processing
<i>ARPC1B</i>	4	R294	R299	R167	Response to estrogen
<i>RPL31</i>	3	R14	R92	R23	Translational initiation
<i>DNAJC8</i>	8	R224	R216	R190	mRNA processing
<i>LCP1</i>	5	R3	R530	R141	Actin filament organization
<i>CBX3</i>	3	R108	R91	R171	Chromatin organization
<i>NOP58</i>	3	R401	R404	R443	rRNA processing
<i>KHDRBS1</i>	2	R442	R436	–	mRNA processing
<i>THRAP3</i>	17	R595	R513	R886	mRNA processing
<i>HCLS1</i>	11	R93	R155	R58	Actin filament polymerization
<i>RPL17</i>	6	R18	R69	R23	Translational initiation
<i>NCL</i>	6	R567	R390	R561	Angiogenesis
<i>PCNP</i>	3	R137	R164	R47	Cell cycle
<i>H1-5</i>	1	R57	–	–	Chromatin organization
<i>EDF1</i>	5	R63	R34	R33	Endothelial cell differentiation
<i>XRCC5</i>	4	R178	R368	R640	Activation of innate immune response
<i>DDX3X</i>	7	R121	R46	R75	Cell differentiation
<i>RPL32</i>	6	R27	R130	R22	Translational initiation
<i>FAM107B</i>	3	R43	R35	R28	Alternative splicing
<i>RPL19</i>	16	R16	R38	R117	Translational initiation
<i>PPIA</i>	6	R55	R148	R144	Protein folding
<i>HSP90AA1</i>	7	R620	R647	R464	Axon extension

than the overall stoichiometry of phosphorylation and N-glycosylation (Fig. 2h).

### Cellular and functional classification of targeted proteins

To obtain an overview of the subcellular compartments and cellular functions that citrullinated proteins associate with, we first clustered the identified citrullination sites into a heat map, which revealed two main clusters (Fig. 2i): one centered on baseline modification sites (cluster 1; yellow), and another containing the majority of citrullination sites upon cell stimulation (cluster 2; red). Next, we performed GO term enrichment analysis to extract the over-represented categories in the identified citrullination clusters as compared with the human proteome. GO Cellular Component (GOCC) analysis highlighted that citrullination sites induced upon differentiation into NLCs and subsequent induction of PADI generally localized to the nucleus, particularly to the nucleolus, nucleoplasm and nuclear inclusion bodies (Fig. 2j). Modification of the ribosome was also observed, in alignment with previous observations<sup>7,42</sup>. The observed nuclear localization of citrullinated proteins aligns with the reported cellular localization of PADI4. Citrullination sites that remained unchanged after cellular perturbation (cluster 1) were predominantly associated with proteins located in the cytosol and extracellular exosome. These sites seemed to be catalyzed by baseline PADI enzyme activity, which remained unaffected by the cellular treatments in this study. GO Biological Processes (GOBP) analysis of the two clusters (Extended Data Fig. 2k) indicated an enrichment for protein targeting to ER and glycolysis in cluster 1, whereas cluster 2 demonstrated an enrichment for histone modification and proteins involved in regulation of RNA binding.

Notably, we identified ~500 citrullination sites in HL60 cells (Fig. 2g) that were not visible in western blot analysis (Fig. 1b). This observation suggests the presence of low-level activation of PADI enzymes and underscores the enhanced sensitivity of our proteomic approach when compared to Western blotting. Among these citrullination sites, 257 were consistently identified across all five experimental conditions (Extended Data Fig. 2l), supporting the hypothesis that they might constitute a basal ‘core citrullinome’. The proteins encompassed within this core citrullinome displayed a significant enrichment in functions related to translation and RNA processing, which suggests that these central cellular functions may be regulated by citrullination under unperturbed cellular conditions.

### Characterizing the PADI4-specific citrullinome

We next aimed to identify substrates specifically citrullinated by PADI4, by performing quantitative citrullination MS analysis of cells treated with the PADI4-specific inhibitor GSK484 (ref. 43) (Fig. 3a). To this end, we treated NLCs with GSK484 for 30 min before the calcium-induced PADI4 activation, and compared GSK484-treated cells with their mock-treated counterparts. All experiments were performed in quadruplicate. To elucidate the concentration-dependent effect of the inhibitor, we performed an analysis across three inhibitor concentrations (1  $\mu$ M,

5  $\mu$ M and 20  $\mu$ M); strong differences in citrullination were observed in western blot results (Fig. 3b). For the MS analysis, we switched to the Orbitrap Exploris 480 mass spectrometer<sup>44</sup>, which improved the sensitivity of the MS methodology. As a result, we observed improved sequencing depth in the analyzed samples (Extended Data Fig. 3a), which overall led to a deeper coverage of the citrullinome (Extended Data Fig. 3b) and improved identification of a total of 14,056 citrullination sites (Supplementary Table 3) on 4,008 protein-coding genes (Supplementary Table 4).

Overall, no changes in protein-expression level were observed when cells were treated with GSK484 (Extended Data Fig. 3c). We defined significantly regulated citrullination sites on the basis of both significance and fold change (shown in the volcano plot in Fig. 3c). No global significant reduction in the non-modified peptides from the same target proteins was observed (Extended Data Fig. 3d). Overall, treating cells with GSK484 led to a global reduction in citrullination signal (Fig. 3c) and to a concomitant decrease in citrullination-site stoichiometry proportionally with increasing concentrations of GSK484 (Extended Data Fig. 3e). To investigate whether the citrullination sites induced via calcium activation of PADI4 correlated with the ones inhibited by GSK484, we compared the increase of citrullination observed upon calcium treatment with the corresponding site-specific changes detected upon inhibition with 20  $\mu$ M GSK484 (Fig. 3d). The strong negative correlation between the two data sets confirmed that the majority of citrullination sites induced upon calcium stimulation of NLCs are PADI4 targets. From the volcano plot (Fig. 3c), we found that 4,957 citrullination sites, located on >2,000 target proteins of citrullination mediated by PADI4, were significantly downregulated upon treatment with GSK484. This corresponded to more than 50% of all identified citrullination target proteins, demonstrating that the regulatory scope of PADI4 is much larger than previously thought<sup>27,30</sup>. Next, we assessed the biological processes associated with PADI4-specific citrullination sites and performed GO analysis. From this, we identified an enrichment for regulation of RNA export from the nucleus and translational initiation, which matches the core citrullinome (Extended Data Fig. 3f). Collectively, we identified citrullination sites on many previously reported ACPAs (Supplementary Table 3), including 44 specifically related to RA, on the basis of the AAgAtlas<sup>45</sup> (Table 2).

To explore the dynamic turnover rate of citrullination at individual sites, we next used a GSK484 inhibitor to investigate the kinetics of citrullination at the proteome-wide scale. By evaluating the reduction of citrullination globally and cumulatively at the site level, we found that a GSK484 concentration value of 1–3  $\mu$ M reduced citrullination by 50% (Fig. 3e) (Supplementary Note 2).

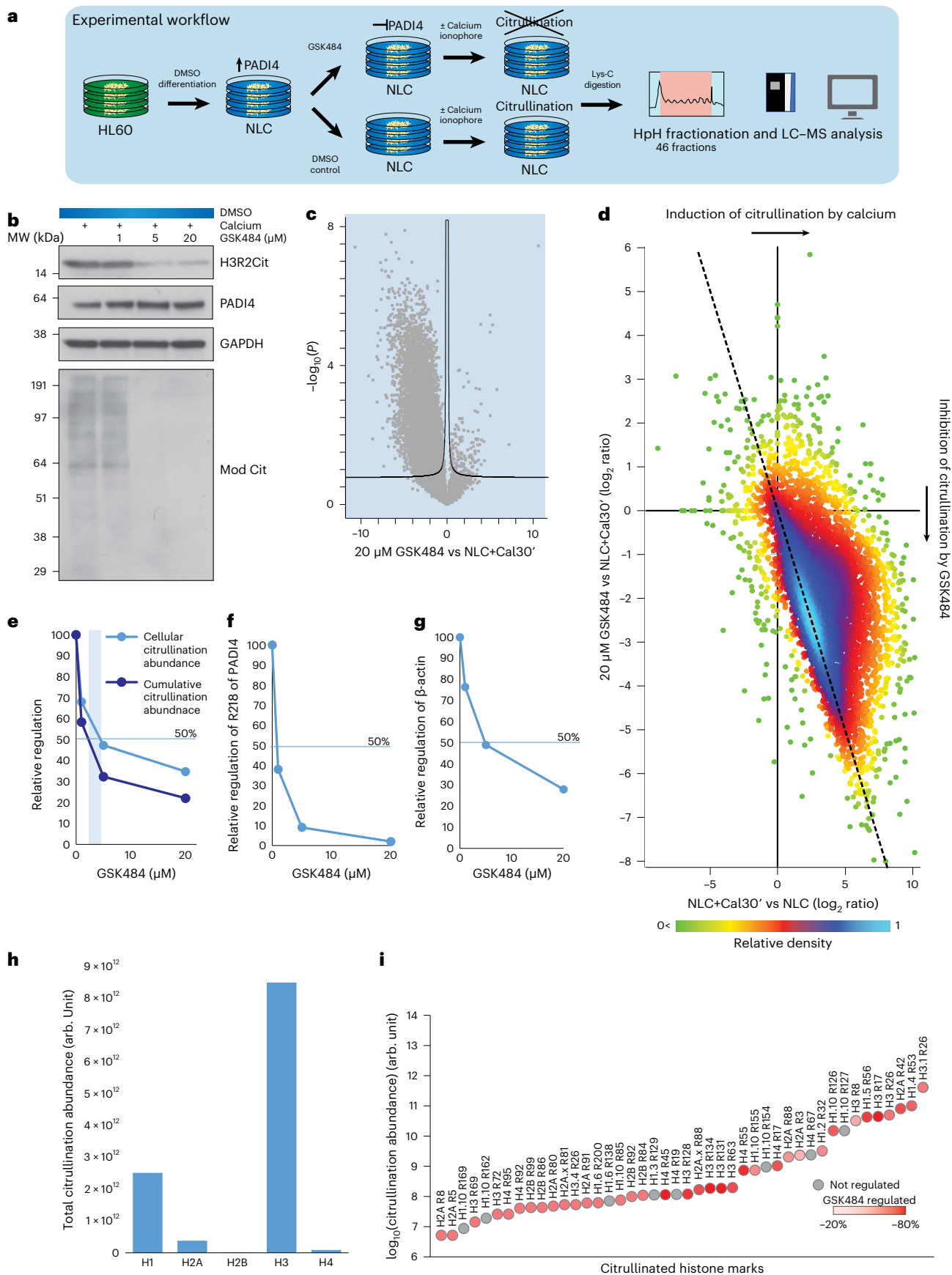
Further investigation revealed that the half-maximal inhibitory concentration ( $IC_{50}$ ) value for GSK484 on the known regulatory autocitrullination of PADI4 (R218) is less than 1  $\mu$ M (Fig. 3f). Looking at additional autocitrullination of PADI4, we found aberrant citrullination events on the less-preferred autocitrullination sites R372 and R374, as previously reported<sup>46</sup>. However, owing to the overall lower abundance

**Fig. 3 | PADI4 inhibition by GSK484.** **a**, Schematic representation of the workflow. HL60 cells were cultured in standard RPMI medium and differentiated by DMSO to neutrophil-like cells (NLCs). Cells were then incubated in Locke’s solution, with varying concentrations (1  $\mu$ M, 5  $\mu$ M, 20  $\mu$ M) of the PADI4 inhibitor GSK484 or a control for 1 h prior to treatment with 4  $\mu$ M calcium ionophore for 30 min. Following Lys-C digestion, the peptides were separated by high-pH fractionation and analyzed using high-resolution LC-MS/MS. Each condition was prepared in quadruplicate. **b**, Western blot analysis of PADI4, H3R2Cit and modified citrulline shows that PADI4 activation by calcium treatment is increasingly inhibited by GSK484. The experiment was repeated three times independently, with similar results. The modified citrulline was run on a different gel than was the GAPDH loading control owing to different sample-preparation requirements. **c**, Volcano plot displaying regulation of sites in response to

treatment with 20  $\mu$ M GSK484. **d**, Density plot of sites detected in response to calcium activation (across *x* axis) and 20  $\mu$ M GSK484 treatment (*y* axis). The dashed line is equivalent to  $y = -x$ . **e**, Regulation of total citrullination abundance and significantly regulated sites in response to increasing GSK484 concentration. The vertical bar represents the concentration range of GSK484 to achieve 50% downregulation of citrullination abundance between total citrullination abundance and the significantly regulated sites. **f**, Regulation of the autocitrullination site R218 of PADI4 in response to increasing GSK484 concentrations relative to control. **g**, Regulation of citrullination on  $\beta$ -actin in response to increasing GSK484 concentrations relative to control. **h**, Total citrullination abundance for histone variants. **i**, Diagram showing citrullination abundance of citrullinated histone marks and their level of regulation in response to 20  $\mu$ M GSK484.

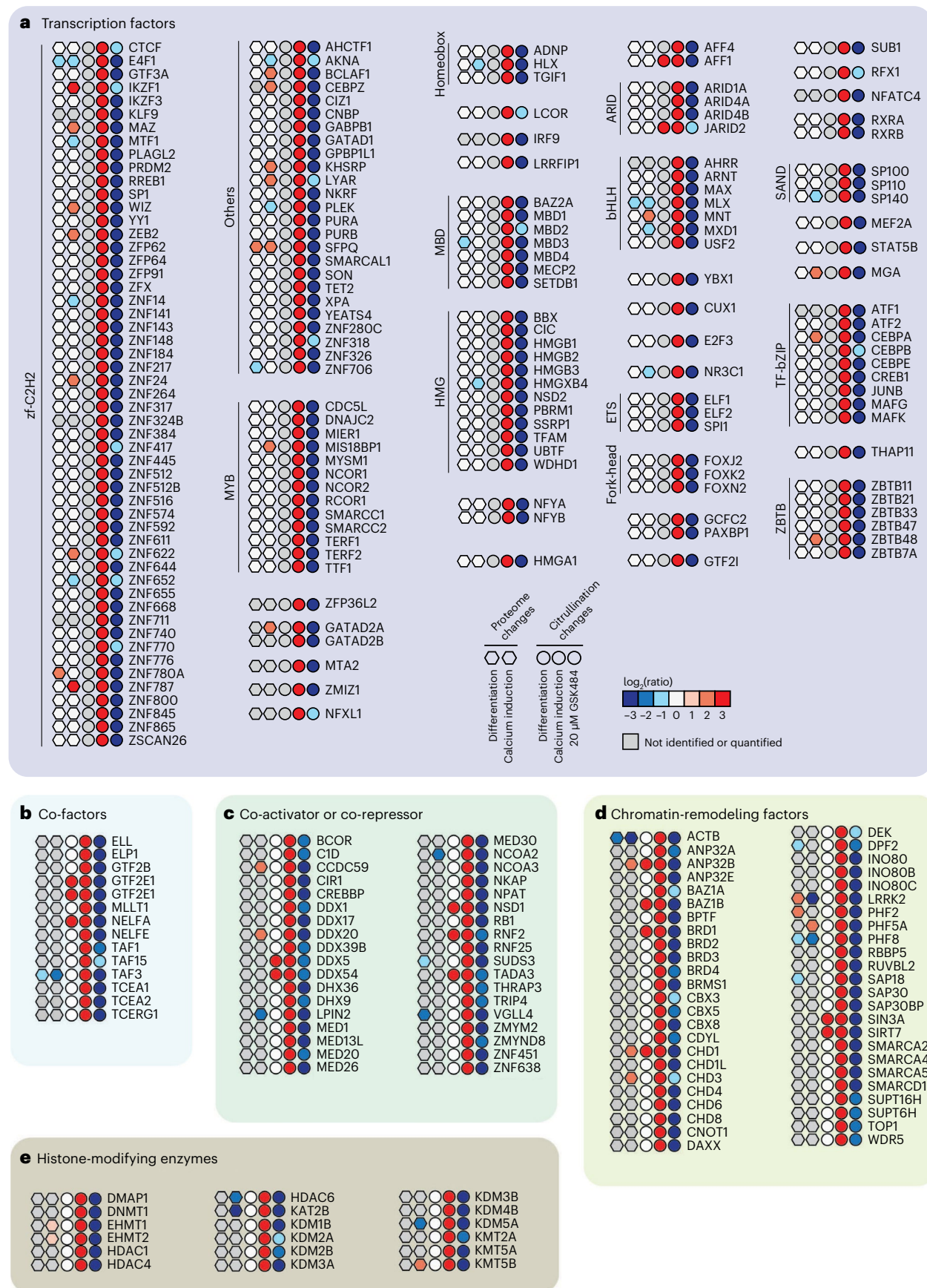
of these modification sites, a proper IC<sub>50</sub> value could not be established across the investigated cell conditions. Still, our data shed light on the regulatory effect that PADI4 inhibitors could have on citrullination of

proteins, including  $\beta$ -actin (Fig. 3g), enolase (Extended Data Fig. 3g) and myelin basic protein (MBP) (Extended Data Fig. 3h)<sup>4,47,48</sup>, that are commonly citrullinated in autoimmune diseases.



**Table 2 | Citrullinated RA autoantigens found in proteomics data set**

Gene name	No. of sites	Protein (Uniprot ID)	Citrullination positions
<i>LMNB1</i>	19	P20700	R10; R14; R51; R67; R136; R197; R220; R221; R226; R234; R250; R258; R315; R336; R397; R402; R407; R410; R413
<i>LMNA</i>	17	P02545	R7; R8; R11; R216; R220; R225; R235; R240; R275; R280; R386; R388; R397; R399; R401; R419; R427; R571; R572
<i>HNRNPA2B1</i>	12	P22626	R12; R21; R54; R82; R95; R99; R147; R153; R185; R325; R350; R352
<i>DEK</i>	11	P35659	R29; R58; R65; R75; R93; R107; R116; R153; R168; R208; R215
<i>HNRNPA1</i>	10	P09651	R31; R47; R75; R88; R92; R97; R122; R140; R146; R178
<i>GTF2F1</i>	10	P35269	R46; R49; R73; R101; R180; R181; R206; R413; R422; R424
<i>TOP1</i>	9	P11387	R17; R138; R140; R163; R210; R362; R364; R693; R708
<i>ENO1</i>	9	P06733	R32; R50; R56; R132; R253; R372; R400; R403; R412
<i>HSPA5</i>	9	P11021	R49; R60; R74; R197; R336; R532; R540; R558; R562
<i>PARP1</i>	8	P09874	R10; R156; R208; R282; R340; R355; R496; R806
<i>MBP</i>	7	P02686	R23; R37; R88; R93; R97; R102; R110
<i>CAST</i>	7	P20810	R86; R206; R269; R301; R310; R378; R610
<i>HMGB2</i>	6	P26583	R10; R24; R70; R73; R110; R177
<i>VIM</i>	6	P08670	R113; R145; R304; R310; R424; R450
<i>PTPN22</i>	5	Q9Y2R2	R357; R458; R510; R672; R756
<i>HMGB1</i>	5	P09429	R70; R73; R110; R163; R24
<i>RAF1</i>	5	P04049	R73; R316; R318; R391; R627
<i>PADI4</i>	4	Q9UM07	R218; R372; R374; R394
<i>HSP90AB1</i>	4	P08238	R291; R506; R604; R612
<i>CAT</i>	4	P04040	R5; R263; R458; R522
<i>SSB</i>	4	P05455	R90; R297; R384; R386
<i>TKT</i>	3	P29401	R205; R471; R474
<i>HMG2</i>	3	P05204	R23; R24; R27
<i>HNRNPD</i>	3	Q14103	R272; R280; R282
<i>CENPB</i>	3	P07199	R66; R588; R593
<i>YBX1</i>	3	P67809	R69; R77; R101
<i>ANXA1</i>	3	P04083	R72; R144; R298
<i>ANXA6</i>	2	P08133	R11; R250
<i>HSPD1</i>	2	P10809	R121; R221
<i>LGALS1</i>	2	P09382	R19; R21
<i>ANXA11</i>	2	P50995	R230; R412
<i>CENPF</i>	2	P49454	R264; R3063
<i>B2M</i>	2	P61769	R32; R101
<i>GPI</i>	2	P06744	R81; R461
<i>TRIM21</i>	2	P19474	R84; R177
<i>ANXA5</i>	1	P08758	R6
<i>SNRPN</i>	1	P63162	R49
<i>CAPNS1</i>	1	P04632	R145
<i>DLAT</i>	1	P10515	R155
<i>CALR</i>	1	P27797	R177
<i>PHB1</i>	1	P35232	R195
<i>RACK1</i>	1	D6R9Z1	R231
<i>PHB2</i>	1	Q99623	R270
<i>NIN</i>	1	Q8N4C6	R1527



**Fig. 4 | Regulation of transcription factors and other factors by PADI4 activity. a–e,** Proteome changes, indicated by color in hexagons, in response to DMSO-induced differentiation and calcium induction, and citrullination changes, indicated by color in circles, in response to differentiation, calcium

induction or 20  $\mu$ M GSK484, for transcription factors (a), co-factors (b), co-activators and co-repressors (c), chromatin-remodeling factors (d) and histone-modifying enzymes (e).

## Histone citrullination

PADI4 is a well-established regulator of gene expression by modifying histone proteins through citrullination<sup>49,50</sup>. Additionally, histone hypercitrullination is known to have a critical role in chromatin decondensation in neutrophils<sup>33</sup>. Hence, our data could offer global insights into PADI4-mediated histone citrullination events.

First, we observed that 12–14% of the total cellular citrullination signal in NLCs localizes to histone proteins (Extended Data Fig. 3i), even when the global citrullination signal is decreased upon GSK484 treatment. By contrast, histones occupy only ~3% of the total protein signal in cells (Extended Data Fig. 3i), demonstrating that core histones are major citrullination targets in human cells.

Overall, we achieved wide sequence coverage of the histone backbones, including the amino-terminal regions (Extended Data Fig. 3j). We identified citrullination sites across all four core histones, with the highest global citrullination signal observed in the case of H3 (Fig. 3h) (Supplementary Table 4). Generally, the histone citrullination signal increased upon calcium stimulation, and conversely was reduced upon PADI4 inhibition, albeit to differing degrees and with different site specificities. The degree of regulation was pronounced for a range of citrullination sites (Fig. 3i), with the most abundant sites detected on H3, H1 and H2A. For example, citrullination of H3R26, H1R53, H2AR42 and H3R17 sites occupied 50%, 13%, 10% and 6%, respectively, of the total histone signal, and GSK484 reduced citrullination at all these sites by >80% (Supplementary Table 3; see Supplementary Note 3 for comment on the modest reduction of citrullination of some markers, such as H3R8, following GSK484 treatment). In our analysis, we quantified nearly all currently known histone citrullination sites while expanding on the current repertoire, enabling us to obtain a comprehensive map of endogenous PADI4-catalyzed histone citrullination marks (Fig. 3i). Although anti-citrullinated H2B antibodies are observed in 90% of people with RA<sup>51</sup>, we found that H2B citrullination occupies <1% of total histone citrullination levels (Fig. 3h), suggesting that even low-abundance histone citrullination marks may constitute an antigenic target of the ACPA immune response. The IC<sub>50</sub> values for H1, H2A and H3 were also estimated (Extended Data Fig. 3k–m).

Collectively, these results provide the first global map of the PADI4-regulated histone sites in mammalian cells, demonstrating that PADI4 targets all four core histones but exhibits varying degrees of regulation, site-specificity and inhibition of citrullination across individual sites and histones. Still, further investigations into the site-specific histone citrullination landscape are likely to enable a mechanistic understanding of PADI4-mediated citrullination and transcriptional regulation.

## PADI4 targets transcriptional regulators

Although PADI4 is a known regulator of gene expression, and a regulatory function of citrullination of non-histone proteins has

been described<sup>52–54</sup>, insight into the site-specific citrullination of transcriptional regulators remains limited. In our data set, we uncovered citrullination of >300 transcription factors, chromatin remodelers, histone-modifying enzymes and transcriptional co-activators (Fig. 4); 179 exhibited increased citrullination upon calcium stimulation and, conversely, decreased citrullination upon calcium stimulation and GSK484 treatment, indicating that citrullination of these proteins is mediated by PADI4. This corroborates the notion that PADI4 acts at the nexus of a broad range of transcription regulatory pathways and supports earlier findings from genome-wide analyses of PADI4 activity that linked the enzyme to active genes and supported its role as a regulator of gene expression<sup>49,50,55</sup>. We further found that PADI4 has a significant preference for targeting DNA-binding regions of the transcriptional regulators (Extended Data Fig. 4a,b and Supplementary Note 4). The downstream targets of the citrullinated transcription factors were investigated (Extended Data Fig. 4c and Supplementary Note 5), providing an example of how our data set can be used to generate hypotheses regarding the potential biological implications of citrullination.

## Properties of the PADI4-dependent citrullinome

Collectively, our large-scale analysis resulted in the high-confidence identification of 14,056 citrullination sites on 4,008 proteins (Fig. 5a), expanding the current knowledge regarding citrullination sites by 16-fold (Fig. 5a). Overall, the citrullination sites identified in this study largely overlap with those found in previous studies (Extended Data Fig. 5a).

To assess the depth of sequencing, we compared our study with a comprehensive study of the human proteome<sup>28</sup>. To this end, we plotted the measured intensity-based absolute quantification (iBAQ) values of the deep human HeLa proteome and aligned these with the corresponding values for the identified citrullinated proteins and the NLC proteome (Fig. 5b). Although we achieved comparable depth of sequencing for our NLC proteome analysis, the citrullinome analysis remained less sensitive, which is not surprising considering the overall stoichiometry of citrullination sites (Fig. 2h). Still, when assessing the depth of sequencing afforded by our data, the detected citrullination sites span seven orders of magnitude in terms of cellular abundance (Fig. 5c), in line with the pattern of other widespread modifications and supporting the excellent sensitivity achieved by our proteomics approach.

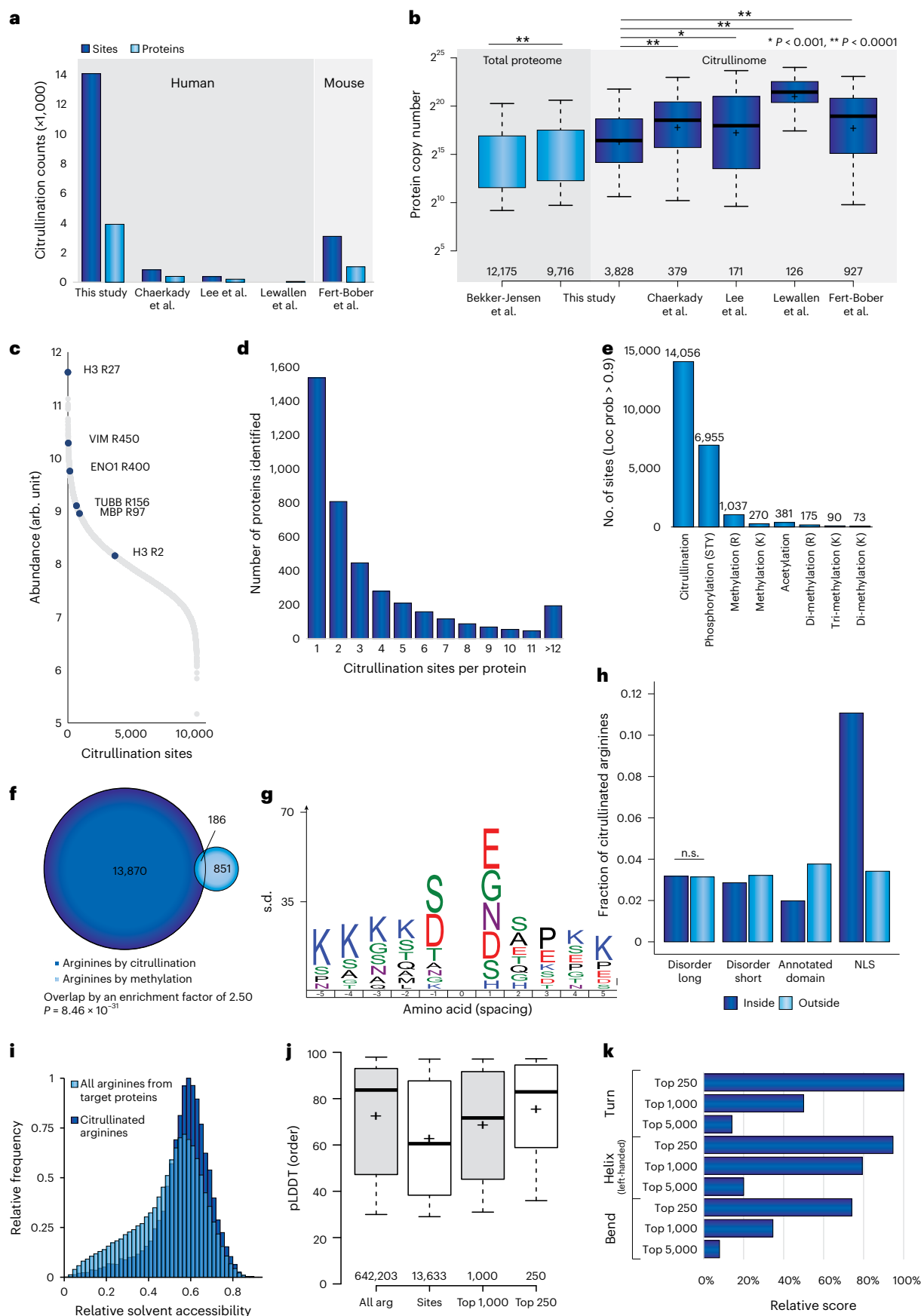
On average, citrullinated proteins harbored 3.4 citrullination sites. Only 1 citrullination site was detected in 40% of proteins (Fig. 5d), 431 proteins harbored >10 citrullination sites and 26 proteins were citrullinated on >20 arginine residues (Supplementary Table 4). The distribution of citrullination across targeted substrates is relatively similar to those of other widespread protein modifications, such as phosphorylation<sup>56</sup>, arginine methylation<sup>57</sup> and SUMOylation<sup>58</sup>.

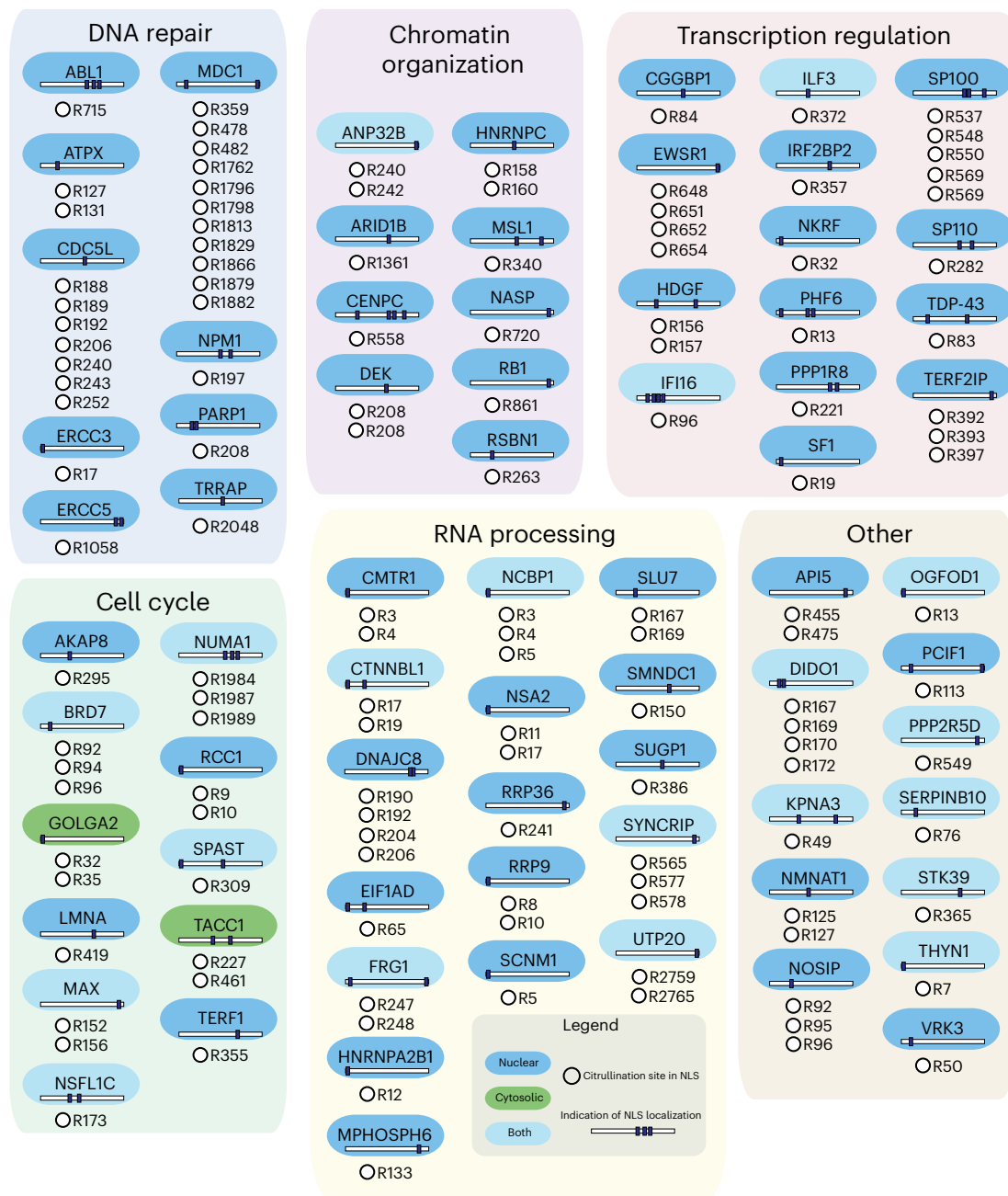
**Fig. 5 | Properties of the citrullinome.** **a**, Number of identified citrullination sites and citrullinated proteins in this study and others<sup>27,30,120,121</sup>. **b**, Protein copy number of proteomes (light blue) and citrullination screens (dark blue) in our study and others<sup>27,30,120,121</sup> relative to deep proteome published by Bekker-Jensen et al.<sup>28</sup>. Whiskers, 95th and 5th percentiles; box limits, third and first quartiles; center bar, median; '+' symbol, average. *n* is listed below the bars. Significance was determined by two-tailed Student's *t* testing, with no correction for multiple comparisons. \**P* < 0.001, \*\**P* < 0.0001. *P* values from left to right,  $3.91 \times 10^{-40}$ ,  $1.97 \times 10^{-101}$ ,  $2.03 \times 10^{-27}$ , 0.00057 and  $3.53 \times 10^{-54}$ . **c**, Distribution of abundance of all sites detected. The most abundant sites on known citrullinated self-antigens are indicated across the distribution. **d**, Number of citrullination sites per protein. **e**, Number of common PTMs detected in our data set. **f**, Pie chart showing the distribution of arginines exclusively targeted by methylation and arginines co-targeted by methylation and citrullination. **g**, IceLogo representation of the enriched amino acid sequence context surrounding identified citrullination sites in the DMSO+Cal30' condition, as detected by Exploris, compared with sequence context of all arginines detected in the proteome from the study. Amino

acid height corresponds to s.d. change. All displayed amino acids represent significant changes, as determined by two-tailed Student's *t*-testing, *n* = 6,245 serine ADP-ribosylation sites, *P* < 0.05. **h**, Analysis of citrullination sites within various protein features reveals a significant enrichment in the localization of citrullination sites with the nuclear localization signal of proteins. **i**, Histogram showing relative solvent accessibility of citrullinated arginines and all arginines from the same target proteins. Two-tailed Spearman nonparametric correlation analysis shows a significant difference ( $7.81 \times 10^{-25}$ ). **j**, Boxplot displaying predicted local distance difference test (pLDDT) order, derived from AlphaFold, of all global arginine residues (All arg), citrullination sites (Sites), the top 1,000 most abundant citrullination sites and the top 250 most abundant citrullination sites. Whiskers, 95th and 5th percentiles; box limits, third and first quartiles; center bar, median; '+' symbol, average. *n* is listed below the bars. **k**, Relative score from the Fisher exact test of different secondary structures from AlphaFold, Turn, Helix left-handed and Bends, showing increased enrichment in response to citrullination abundance.

Citrullination and arginine methylation are described as functionally interacting on histones in the context of epigenetics and regulation of splicing factors<sup>59,60</sup>. However, little is known as to whether the two

modifications occupy the same arginine residues within proteins. To assess this, we first analyzed our data for peptides modified with various classes of PTMs, such as phosphorylation, acetylation and various





**Fig. 6 | Targeting of nuclear localization signals.** Citrullination site number overview. Proteins have been divided into functional families: DNA repair (top left), chromatin organization (top middle), transcription regulation (top right), cell cycle (bottom left), RNA processing (bottom middle) and other (bottom right).

lysine and arginine methylation isoforms (Fig. 5e). In total, we were able to find 6,955 phosphorylation, 381 acetylation and 1,645 methylation forms, of which the majority (1,037) were identified as arginine mono-methylation sites. Notably, 186 citrullination sites overlapped site-specifically with identified arginine mono-methylation sites, a significantly higher degree of overlap than what is expected by chance ( $P < 3 \times 10^{-107}$ ) (Fig. 5f). Thus, we show that citrullination and arginine mono-methylation show a statistically significant overlap on the same arginine residues on a proteome-wide scale.

Next, we identified a mixed sequence motif for targeting by PADI4 (Fig. 5g), including aspartic acid and serine at position -1 and aspartic acid and glycine at +1, as previously reported<sup>30</sup>, alongside other amino acids at +1 and -1. The motif is not as strong as detected for other PTMs; however, this may be due to differences in detection strategies.

Considering the large number of identified citrullination sites, we aimed to test whether citrullination sites localize to specific domains or structural regions of the modified proteins. We found that there is an enrichment for citrullination outside short disordered regions and annotated domains (Fig. 5h), which matches a previous report of citrullination preferentially targeting intrinsically disordered protein regions<sup>3</sup>. Additionally, we identified a strong enrichment for citrullination inside the nuclear localization signal (NLS) of proteins (Fig. 5h), along with a predominant targeting of factors related to nuclear processes, including DNA damage, chromatin organization, transcriptional regulation and the cell cycle (Fig. 6). This supports the notion that citrullination has an unappreciated role in regulating nuclear shuttling and protein localization. Consistent with this notion, we found that citrullination sites of PARP1 (R208), NPM1 (R197) and

TDP-43 (R83) are targeted to arginine residues within their respective NLS domains, which are known to regulate the cytoplasmic translocation of these proteins<sup>61–63</sup>. We further investigated the relative solvent accessibility (RSA) of the citrullination events, to explore whether citrullination primarily occurs on buried or exposed residues. We found that citrullination occurs on exposed arginine residues, evidenced by a significantly higher RSA among the arginine residues targeted for citrullination than for all other arginine residues in the targeted proteins<sup>64</sup> (Fig. 5i;  $P = 7.81 \times 10^{-25}$ ). This is supported by a significantly lower number of nearby alpha-carbon atoms for the citrullinated arginine residues than for all arginine residues in the targeted proteins (Extended Data Fig. 5b;  $P = 7.53 \times 10^{-32}$ ).

Finally, we aligned the citrullination sites with the AlphaFold database and, on the basis of the local distance difference test, observed that sites tended to localize within lower-order regions and primarily disordered regions, as compared with arginine residues, across all citrullinated proteins. However, upon considering more abundant citrullination sites, we observed an increase in structural order as the abundance increased (Fig. 5j). Furthermore, we investigated which secondary structures are preferentially targeted by PADI4 and found enrichment of bend, left-handed helices and turns among the most abundant citrullination events (Fig. 5k), whereas citrullination globally targets disordered regions (Fig. 5j). Collectively, citrullination preferentially targets exposed arginine residues and disordered regions on a global scale. However, the most abundant citrullination sites occur in regions of increased structural order.

### Validation of citrullination sites by peptide microarray

To highlight the value of our proteomics screen, we synthesized all identified citrullinated sites onto peptide microarray chips and assessed whether they were recognized by ACPAs from people with RA.

To this end, we obtained synovial fluid samples from people with RA who were either negative or positive for anti-CCP antibodies, a proxy for ACPAs, and investigated their capacity to interact with the citrullinated sites we identified in our MS screen (Fig. 7a). Synovial fluid from individuals with ankylosing spondylitis was used as a non-RA control. In addition to the identified citrullination sites, our microarray chip featured peptide sequences encompassing the multi-citrullinated peptides we identified, their corresponding non-modified peptide variants, random arginine residues derived from the same proteins as the identified sites, peptides known to be citrullinated in the literature<sup>65–67</sup> and a commercially available PEPperPRINT CCP array (linearized) (Fig. 7b). In total, we synthesized duplicates of 32,653 peptide sequences on the microarray chip, resulting in a total of 65,306 sequences. The intensity derived from antibody binding to these citrullinated peptides was compared with that of the binding to their unmodified (baseline) variants on the same chip.

Using this array, we detected autoantibodies in all three groups. As expected, only in the anti-CCP-positive RA group did the level of binding to the citrullinated form of the antigens exceed that of the binding to the baseline peptides (Fig. 7c). We noted that there were low background signals from the secondary antibody control, suggesting that the binding of autoantibodies was antigen-specific in all cases. Additionally, in the raw images used for quantification (Extended Data Fig. 6a), we noted that while the control samples do show some binding of antibodies, the binding is stronger in the wells containing synovial fluid from people with RA, in particular in those containing synovial fluid from those who are the anti-CCP-positive.

Upon closer examination, we noticed that the controls and the anti-CCP-negative RA group overall demonstrated increased antibody reactivity to the unmodified (baseline) peptides, whereas the anti-CCP-positive RA group exhibited stronger reactivity, particularly towards citrullinated peptides (Fig. 7c). This observation supports the notion that the ACPA binding in the anti-CCP-positive group is more specific towards the citrullinated sites identified in our MS screen.

Next, we investigated the potential impact of cellular abundance of citrullination sites, as detected in our MS screen, on overall ACPA binding. To achieve this, we correlated the peptide signal intensities obtained from our proteomics data with the ACPA binding levels observed in the array data. This analysis revealed a notably heightened response directed at citrullination sites with substantial cellular abundance (Fig. 7d). Furthermore, we observed a pronounced ACPA binding towards multi-citrullinated peptides, indicating that patches of citrullination in close proximity on proteins might elicit the strongest binding. Given that the existence of multiple citrulline sites within a single peptide could facilitate several simultaneous binding events from different ACPAs, we investigated the spacing between the sites to see if such a model would be permissive. Interestingly, the preferred spacing between citrullines on the peptide is short; just one amino acid separates the citrullinated residues for the most reactive peptides (Extended Data Fig. 6b), which makes simultaneous binding unlikely. This suggests that at least some ACPAs could have the capacity to recognize multi-citrullinated epitopes.

We then investigated the sequence motif for the strongest interactors in the two RA groups and found a strong preferential binding of ACPAs from the anti-CCP-positive RA group towards a Cit-Gly motif (Fig. 7e), which correlates with previously reported sequence motifs for ACPA recognition<sup>65,66,68,69</sup>. In addition to the Cit-Gly motif, we also found that positions –3 to –1 upstream of the Cit residue were enriched for bulky aliphatic amino acids, which create a hydrophobic patch that has not previously been described. In combination, these features underpin a recognition pattern that drives the specific ACPA binding for this anti-CCP-positive subgroup. Conversely, we observed an under-representation of glycine at the +1 position, and

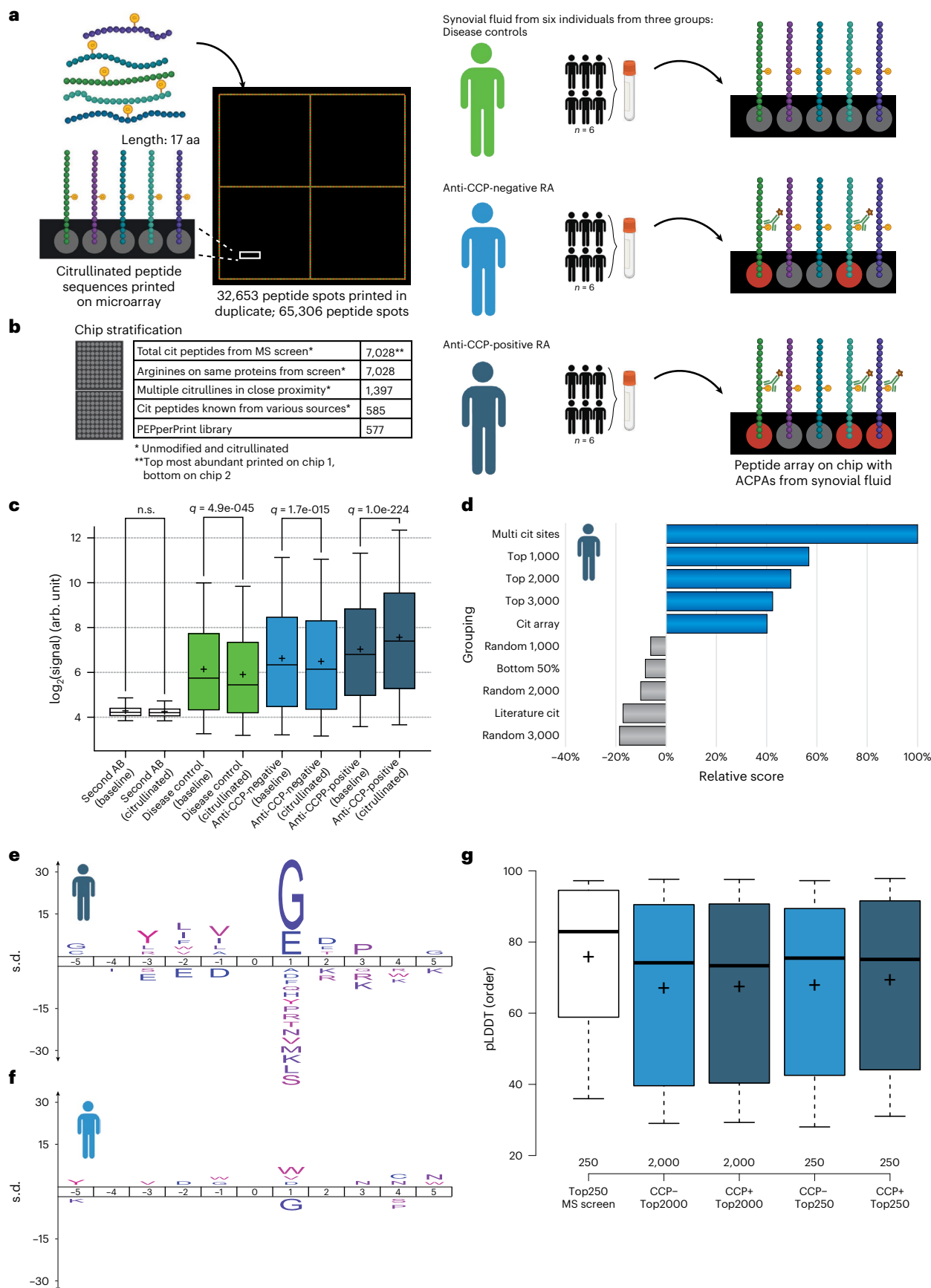
### Fig. 7 | Validation of mapped citrullination sites by peptide microarray.

**a**, Representation of workflow. Citrullination sites, detected by the MS screen, were printed on a microarray chip with the citrullination site centered in a 17-amino-acid peptide. Identical microarray chips with the citrullinated peptides and controls were then exposed to synovial fluids from three groups: (1) people with ankylosing spondylitis, (2) people with RA who were negative for anti-CCP antibodies and (3) people with RA who were positive for anti-CCP antibodies. Fluids were combined from six individuals in each group. The binding of the ACPAs to the citrullinated and control peptides was analyzed. **b**, Overview of peptides on the microarray. Owing to the limited capacity, the two first categories were split onto two chips. The other controls were included on both chips. **c**, Within-chip analysis of antibody binding in response to unmodified (baseline) peptide or citrullinated peptide. Second AB, secondary antibody binding. Whiskers, 95th and 5th percentiles; box limits, third and first quartiles; center bar, median; '+' symbol, average.  $n = 65,306$  peptide spots. **d**, Relative score from the Fisher exact testing of the antibody-binding response to multiple

citrullinated sites (Multi cit sites); the top 1,000, 2,000 and 3,000 most abundant sites from the MS screen; PEPperCHIP cyclic citrullinated peptides (Cit array); 1,000 random sites; the bottom 50% responsive sites; 2,000 random sites; sites from the literature (Literature cit); and 3,000 random sites. **e**, Sequence motif of 1,000 singly modified citrullination sites mapped in the MS screen that overlap with best reactive sites from the anti-CCP-positive RA group; background is all other sites from the MS screen. **f**, Sequence motif of 1,000 singly modified citrullination sites mapped in the MS screen overlapping with reactive sites from the anti-CCP-negative RA group, compared to all other sites. **g**, Boxplot displaying pLDDT order, derived from AlphaFold, of the top 250 most abundant citrullination sites from the MS screen as a reference, the top 2,000 most reactive sites from the anti-CCP-negative RA group (CCP-), the top 2,000 most reactive sites from the anti-CCP-positive RA group (CCP+), the top 250 most reactive sites from the anti-CCP-negative RA group and the top 250 most reactive sites from the anti-CCP-positive RA group. Whiskers, 95th and 5th percentiles; box limits, third and first quartiles; center bar, median; '+' symbol, average.  $n$  is listed below bars.

no preference for hydrophobic amino acids in the anti-CCP-negative RA group (Fig. 7f). This observation raises the possibility that a distinct antibody-specificity profile might underlie the lack of response

to conventional anti-CCP tests in this group. Following the sequence motif analysis, we explored whether the most reactive sites in the anti-CCP-negative and anti-CCP-positive RA groups reside within



regions of defined secondary structure, as was the case for the increasingly abundant citrullination sites during the MS screen (Fig. 5i). This was not the case for the top 2,000 and top 250 reactive peptides in the two groups, as the most reactive peptides target lower-order regions (Fig. 7g).

In conclusion, these results underscore the biological importance and potential clinical relevance of the identified citrullination sites, and their value as a resource to the field. We find that antibodies from people with RA that are classified as either anti-CCP-positive or anti-CCP-negative show a distinct preference for different citrullination motifs, and that multiple citrullination events that form a patch capture most ACPAs. We are also able to demonstrate that the ACPA binding for the anti-CCP-positive RA group is specific for the identified sites over baseline peptides.

## Discussion

We present a high-confidence atlas focusing on the PADI4-regulated citrullinome, characterizing more than 14,000 sites using high-accuracy mass spectrometry in the HL60 model system. Our findings expand current knowledge by 16-fold, offering insights into the biological implications of citrullination. Pinpointing citrullination sites across the proteome enhances the understanding of functional consequences, aiding the development of reagents for biochemical and cell-biology studies. Using the PADI4 inhibitor GSK488, we demonstrate widespread regulation of citrullination by PADI4, suggesting that the scope of PADI4 targeting is broader than anticipated. Our observations propose a paradigm of PADI4-mediated citrullination as a high-density ‘citrullination spray’ across various substrates, akin to other modifications such as SUMOylation and ADP-ribosylation<sup>70</sup>.

Considering the widespread interest in targeting PADI enzymes to treat a variety of human pathologies<sup>71,72</sup>, evaluation of the citrullinated proteome under treatment regimens with PADI-specific drugs, as demonstrated in this study, provides insightful information and reveals useful endogenous biomarkers to interrogate pathway function. For example, increasing evidence suggests that citrullinated versions of endogenous proteins constitute autoantigens in a variety of autoimmune disorders, and that corresponding ACPAs could serve as diagnostic and prognostic markers<sup>73,74</sup>. In support of this, our data detail the specific citrullination sites related to 44 autoantigens specific to RA.

Our data also provide insights into the specificity of PADI4 as a peptidylarginine deiminase, and we find that PADI4 citrullinates substrate proteins independently of linear sequence. We further find that citrullination is primarily directed to disordered regions, which supports a role for the process in modulating protein binding, considering that disordered protein regions hold central roles in protein interaction networks. Similarly, we find that citrullination localizes outside annotated domains, with the exception of the strong enrichment of citrullination sites inside the nuclear localization signal of proteins. Intriguingly, citrullination preferentially targets exposed arginine residues in disordered regions globally, but the most abundant sites occur in regions of increased structural order. GO analysis of the genes regulated by citrullinated transcription factors shows enrichment for pathways involved in immune responses and terms related to the maintenance of the skin barrier, areas in which physiological citrullination is known to have a direct role<sup>33,75-77</sup>. These results not only offer indirect insights, but also highlight the rich information within our data, suggesting that exploration of the functional role of citrullination beyond individual sites could be fruitful. Although in-depth studies are needed to comprehend PADI4’s regulatory role in gene regulation, our research contributes to a systems-wide understanding of the connection between citrullination and gene transcription. Additionally, we reveal a notable overlap in arginine targeting by citrullination and methylation, supporting previous indications of a potential inhibitory interplay between these two post-translational modifications, both recognized for their roles in regulating transcriptional activity.

Histone modifications are widely linked to transcriptional regulation, and great efforts have been made to map the genomic regions targeted by various histone modification marks<sup>78</sup>. In regard to citrullination, however, the understanding of the enzymatic specificity and dynamics remains limited<sup>79</sup>. Within our MS-based quantitative atlas, we quantify histone citrullination sites exclusively catalyzed by PADI4, which could help to more accurately define genomic regions that are occupied by catalytically active PADI4, especially considering that epigenetic mapping of histone marks typically focuses on histones H3 and H4.

Our kinetic analyses of PADI4 inhibition provide the first proteome-wide survey of site-specific citrullination events directly mediated by PADI4 through identification of a large set of sites that are downregulated upon treatment with GSK484. This regulation is reminiscent of the inhibition of other enzymatic regulations within, for example, phosphorylation and acetylation signaling<sup>80,81</sup>. Our results demonstrate that PADI4-mediated citrullination signaling is regulated to a similar degree as other widespread modifications.

To demonstrate the utility of our citrulline atlas for biological insights, we investigated how synovial fluid antibodies from people with RA react to the identified sites, given the autoimmune response directed at citrullinated antigens in RA, especially in people with anti-CCP antibodies<sup>21,82,83</sup>. To this end, we synthesized all identified citrullination sites as linear peptides on a microarray. Upon testing whether the sites were recognized by autoantibodies in synovial fluid samples, we observed binding in all groups; the anti-CCP-positive group exhibited stronger binding, consistent with clinical severity<sup>84-88</sup>. The decreased antibody binding towards citrullinated peptides in the disease control and anti-CCP-negative groups (Fig. 7c) might be due to disruption of native binding epitopes lacking citrulline. Although our proteomics data did not reveal a defined sequence motif, our microarray data unveiled a recognition motif in anti-CCP-positive individuals, a hydrophobic patch followed by Cit-Gly that has not previously been described<sup>66,68</sup>. This suggests that PADI4’s catalytic activity is not directed toward a specific motif (Fig. 5g) but might instead act as a ‘citrullination spray’. Still, antibodies from anti-CCP-negative individuals with RA showed no preference towards Cit-Gly motifs, potentially explaining the challenges in their identification using contemporary anti-CCP tests biased toward detection of Cit-Gly motifs<sup>89</sup>.

Hence, we propose that the new motif identified in this study, [3Φ-Cit-Gly], where 3Φ denotes three bulky hydrophobic residues, may enhance RA diagnosis specificity. Additionally, the over-representation of multi-citrullinated peptides for binding suggests that the increased density of citrullination in a concentrated patch may improve the performance of the next-generation ACPA test for enhanced RA diagnosis.

In summary, our citrulline site atlas enhances our understanding of RA and could help improve RA diagnostic techniques. Citrulline is integral to several autoimmune disorders, including multiple sclerosis, lupus, psoriasis and inflammatory bowel disease<sup>21</sup>.

Unlike RA, in which ACPA reactivity is central, specific autoantibodies for these diseases are yet to be discovered. Therefore, our resource data could improve early diagnosis for individuals with autoimmune diseases by facilitating the discovery of new biomarkers.

In conclusion, this in-depth analysis provides a powerful resource for identifying and quantifying citrullinated residues induced by PADI4. It establishes a framework for decoding PADI4 functions in various biological processes and enhances our understanding of PADI4-specific inhibitors. This comprehensive resource allows for easy future investigation of individual site utilization.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41594-024-01214-9>.

## References

- Vossenaar, E. R., Zendman, A. J. W., van Venrooij, W. J. & Pruijn, G. J. M. PAD, a growing family of citrullinating enzymes: genes, features and involvement in disease. *BioEssays* **25**, 1106–1118 (2003).
- Christophorou, M. A. The virtues and vices of protein citrullination. *R. Soc. Open Sci.* **9**, 220125 (2022).
- Tarcsa, E. et al. Protein unfolding by peptidylarginine deiminase. Substrate specificity and structural relationships of the natural substrates trichohyalin and filaggrin. *J. Biol. Chem.* **271**, 30709–30716 (1996).
- Musse, A. A. et al. Peptidylarginine deiminase 2 (PAD2) overexpression in transgenic mice leads to myelin loss in the central nervous system. *Dis. Models Mech.* **1**, 229–240 (2008).
- Christophorou, M. A. et al. Citrullination regulates pluripotency and histone H1 binding to chromatin. *Nature* **507**, 104–108 (2014).
- Deplus, R. et al. Citrullination of DNMT3A by PADI4 regulates its stability and controls DNA methylation. *Nucleic Acids Res.* **42**, 8285–8296 (2014).
- Falcão, A. M. et al. PAD2-mediated citrullination contributes to efficient oligodendrocyte differentiation and myelination. *Cell Rep.* **27**, 1090–1102 (2019).
- Méchin, M. C., Takahara, H. & Simon, M. Deimination and peptidylarginine deiminases in skin physiology and diseases. *Int. J. Mol. Sci.* **21**, 566 (2020).
- Zhang, X. et al. Peptidylarginine deiminase 1-catalyzed histone citrullination is essential for early embryo development OPEN. *Sci. Rep.* **6**, 38727 (2016).
- Kholia, S. et al. A novel role for peptidylarginine deiminases in microvesicle release reveals therapeutic potential of PAD inhibition in sensitizing prostate cancer cells to chemotherapy. *J. Extracell. Vesicles* **4**, 26192 (2015).
- Haider, L. et al. The topography of demyelination and neurodegeneration in the multiple sclerosis brain. *Brain* **139**, 807–815 (2016).
- Sokolove, J. et al. Citrullination within the atherosclerotic plaque: a potential target for the anti-citrullinated protein antibody response in rheumatoid arthritis. *Arthritis Rheum.* **65**, 1719–1724 (2013).
- Khandpur, R. et al. NETs are a source of citrullinated autoantigens and stimulate inflammatory responses in rheumatoid arthritis. *Sci. Transl. Med.* **5**, 178ra40 (2013).
- Corsiero, E., Pratesi, F., Prediletto, E., Bombardieri, M. & Migliorini, P. NETosis as source of autoantigens in rheumatoid arthritis. *Front. Immunol.* **7**, 485 (2016).
- Singh, U., Singh, S., Singh, N. K., Verma, P. K. & Singh, S. Anticyclic citrullinated peptide autoantibodies in systemic lupus erythematosus. *Rheumatol. Int.* **31**, 765–767 (2011).
- Gupta, S. & Kaplan, M. J. The role of neutrophils and NETosis in autoimmune and renal diseases. *Nat. Rev. Nephrol.* **12**, 402–413 (2016).
- Suzuki, A. et al. Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nat. Genet.* **34**, 395–402 (2003).
- Yamada, R., Suzuki, A., Chang, X. & Yamamoto, K. Peptidylarginine deiminase type 4: identification of a rheumatoid arthritis-susceptible gene. *Trends Mol. Med.* **9**, 503–508 (2003).
- Lee, Y. H. & Bae, S. C. Association between susceptibility to rheumatoid arthritis and PADI4 polymorphisms: a meta-analysis. *Clin. Rheumatol.* **35**, 961–971 (2016).
- Schellekens, G. A., de Jong, B. A., van den Hoogen, F. H., van de Putte, L. B. & van Venrooij, W. J. Citrulline is an essential constituent of antigenic determinants recognized by rheumatoid arthritis-specific autoantibodies. *J. Immunol.* **195**, 8–16 (1998).
- Ciesielski, O. et al. Citrullination in the pathology of inflammatory and autoimmune disorders: recent advances and future perspectives. *Cell. Mol. Life Sci.* **79**, 94 (2022).
- Hensen, S. M. M. & Pruijn, G. J. M. Methods for the detection of peptidylarginine deiminase (PAD) activity and protein citrullination. *Mol. Cell. Proteom.* **13**, 388–396 (2014).
- Slade, D. J., Subramanian, V., Fuhrmann, J. & Thompson, P. R. Chemical and biological methods to detect post-translational modifications of arginine. *Biopolymers* **101**, 133–143 (2014).
- Senshu, T., Sato, T., Inoue, T., Akiyama, K. & Asaga, H. Detection of citrulline residues in deiminated proteins on polyvinylidene difluoride membrane. *Anal. Biochem.* **203**, 94–100 (1992).
- Tutturen, A. E. V., Fleckenstein, B. & De Souza, G. A. Assessing the citrullinome in rheumatoid arthritis synovial fluid with and without enrichment of citrullinated peptides. *J. Proteome Res.* **13**, 2867–2873 (2014).
- De Ceuleneer, M. et al. Modification of citrulline residues with 2,3-butanedione facilitates their detection by liquid chromatography/mass spectrometry. *Rapid Commun. Mass Spectrom.* **25**, 1536–1542 (2011).
- Fert-Bober, J. et al. Mapping citrullinated sites in multiple organs of mice using hypercitrullinated library. *J. Proteome Res.* **18**, 2270–2278 (2019).
- Bekker-Jensen, D. B. et al. An optimized shotgun strategy for the rapid generation of comprehensive human proteomes. *Cell Syst.* **4**, 587–599 (2017).
- Nakashima, K. et al. Molecular characterization of peptidylarginine deiminase in HL-60 cells induced by retinoic acid and 1 $\alpha$ ,25-dihydroxyvitamin D<sub>3</sub>. *J. Biol. Chem.* **274**, 27786–27792 (1999).
- Lee, C.-Y. et al. Mining the human tissue proteome for protein citrullination. *Mol. Cell. Proteom.* **17**, 1378–1391 (2018).
- Kelstrup, C. D., Young, C., Lavalley, R., Nielsen, M. L. & Olsen, J. V. Optimized fast and sensitive acquisition methods for shotgun proteomics on a quadrupole orbitrap mass spectrometer. *J. Proteome Res.* **11**, 3487–3497 (2012).
- Alyami, H. M. et al. Role of NOD1/NOD2 receptors in *Fusobacterium nucleatum* mediated NETosis. *Microb. Pathogenesis* **131**, 53–64 (2019).
- Wang, Y. et al. Histone hypercitrullination mediates chromatin decondensation and neutrophil extracellular trap formation. *J. Cell Biol.* **184**, 205–213 (2009).
- Kelstrup, C. D. et al. Rapid and deep proteomes by faster sequencing on a benchtop quadrupole ultra-high-field orbitrap mass spectrometer. *J. Proteome Res.* **13**, 6187–6195 (2014).
- Zhou, Y. et al. Spontaneous secretion of the citrullination enzyme PAD2 and cell surface exposure of PAD4 by neutrophils. *Front. Immunol.* **8**, 1200–1200 (2017).
- Foulquier, C. et al. Peptidyl arginine deiminase type 2 (PAD-2) and PAD-4 but not PAD-1, PAD-3, and PAD-6 are expressed in rheumatoid arthritis synovium in close association with tissue inflammation. *Arthritis Rheum.* **56**, 3541–3553 (2007).
- Lakschevitz, F. S. et al. Identification of neutrophil surface marker changes in health and inflammation using high-throughput screening flow cytometry. *Exp. Cell. Res.* **342**, 200–209 (2016).
- Eldewi, D. M. et al. Expression levels of complement regulatory proteins (CD35, CD55 and CD59) on peripheral blood cells of patients with chronic kidney disease. *Int. J. Gen. Med.* **12**, 343–351 (2019).
- Chick, J. M. et al. An ultra-tolerant database search reveals that a myriad of modified peptides contributes to unassigned spectra in shotgun proteomics HHS Public Access Author manuscript. *Nat. Biotechnol.* **33**, 743–749 (2015).

40. Hao, G. et al. Neutral loss of isocyanic acid in peptide CID Spectra: a novel diagnostic marker for mass spectrometric identification of protein citrullination. *J. Am. Soc. Mass Spectrom.* **20**, 723–727 (2008).
41. Prus, G., Hoegl, A., Weinert, B. T. & Choudhary, C. Analysis and interpretation of protein post-translational modification site stoichiometry. *Trends Biochem. Sci.* **44**, 943–960 (2019).
42. Guo, Q., Bedford, M. T. & Fast, W. Discovery of peptidylarginine deiminase-4 substrates by protein array: antagonistic citrullination and methylation of human ribosomal protein S2. *Mol. Biosyst.* **7**, 2286–2295 (2011).
43. Mondal, S. & Thompson, P. R. Protein arginine deiminases (PADs): biochemistry and chemical biology of protein citrullination. *Acc. Chem. Res.* **52**, 818–832 (2019).
44. Bekker-Jensen, D. B. et al. A compact quadrupole-orbitrap mass spectrometer with FAIMS interface improves proteome coverage in short LC gradients. *Mol. Cell. Proteom.* **19**, 716–729 (2020).
45. Wang, D. et al. AAgAtlas 1.0: a human autoantigen database. *Nucleic Acids Res.* **45**, D769–D776 (2017).
46. Mondal, S. et al. Site-specific incorporation of citrulline into proteins in mammalian cells. *Nat. Commun.* **12**, 45 (2021).
47. Kinloch, A. et al. Identification of citrullinated alpha-enolase as a candidate autoantigen in rheumatoid arthritis. *Arthritis Res. Ther.* **7**, 1421–1429 (2005).
48. Olsson, T. et al. Increased numbers of T cells recognizing multiple myelin basic protein epitopes in multiple sclerosis. *Eur. J. Immunol.* **22**, 1083–1087 (1992).
49. Cuthbert, G. L. et al. Histone deimination antagonizes arginine methylation. *Cell* **118**, 545–553 (2004).
50. Wang, Y. et al. Human PAD4 regulates histone arginine methylation levels via demethylation. *Science* **306**, 279–283 (2004).
51. Sohn, D. H. et al. Local joint inflammation and histone citrullination in a murine model of the transition from preclinical autoimmunity to inflammatory arthritis. *Arthritis Rheumatol.* **67**, 2877–2887 (2015).
52. Lee, Y.-H., Coonrod, S. A., Kraus, W. L., Jelinek, M. A. & Stallcup, M. R. Regulation of coactivator complex assembly and function by protein arginine methylation and demethylation. *Proc. Natl Acad. Sci. USA* **102**, 3611–3616 (2005).
53. Wang, S. & Wang, Y. Peptidylarginine deiminases in citrullination, gene regulation, health and pathogenesis. *Biochim. Biophys. Acta* **1829**, 1126–1135 (2013).
54. Sharma, P. et al. Arginine citrullination at the C-terminal domain controls RNA polymerase II transcription. *Mol. Cell* **73**, 84–96 (2019).
55. Zhai, Q., Wang, L., Zhao, P. & Li, T. Role of citrullination modification catalyzed by peptidylarginine deiminase 4 in gene transcriptional regulation. *Acta Biochim. Biophys. Sin.* **49**, 567–572 (2017).
56. Huttlin, E. L. et al. A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **143**, 1174–1189 (2010).
57. Larsen, S. C. et al. Proteome-wide analysis of arginine monomethylation reveals widespread occurrence in human cells. *Sci. Signal.* **9**, rs9 (2016).
58. Hendriks, I. A. et al. Site-specific mapping of the human SUMO proteome reveals co-modification with phosphorylation. *Nat. Struct. Mol. Biol.* **24**, 325–336 (2017).
59. Fuhrmann, J. & Thompson, P. R. Protein arginine methylation and citrullination in epigenetic regulation. *ACS Chem. Biol.* **11**, 654–668 (2016).
60. Clancy, K. W. et al. Citrullination/methylation crosstalk on histone H3 regulates ER-target gene transcription. *ACS Chem. Biol.* **12**, 1691–1702 (2017).
61. Tanikawa, C. et al. Regulation of protein citrullination through p53/PADI4 network in DNA damage response. *Cancer Res.* **69**, 8761–8769 (2009).
62. Schreiber, V., Molinete, M., Boeuf, H., de Murcia, G. & Ménissier-de Murcia, J. The human poly(ADP-ribose) polymerase nuclear localization signal is a bipartite element functionally separate from DNA binding and catalytic activity. *EMBO J.* **11**, 3263–3269 (1992).
63. Doll, S. G. et al. Recognition of the TDP-43 nuclear localization signal by importin  $\alpha$ 1/ $\beta$ . *Cell Rep.* **39**, 111007 (2022).
64. Hoie, M. H. et al. NetSurfP-3.0: accurate and fast prediction of protein structural features by protein language models and deep learning. *Nucleic Acids Res.* **50**, W510–W515 (2022).
65. Snir, O. et al. Multiple antibody reactivities to citrullinated antigens in sera from patients with rheumatoid arthritis: association with HLA-DRB1 alleles. *Ann. Rheum. Dis.* **68**, 736–743 (2009).
66. Trier, N. H., Dam, C. E., Olsen, D. T., Hansen, P. R. & Houen, G. Contribution of peptide backbone to anti-citrullinated peptide antibody reactivity. *PLoS ONE* **10**, e0144707 (2015).
67. Park, M., Pyun, J. C., Akter, H., Nguyen, B. T. & Kang, M. J. Evaluation of a specific diagnostic marker for rheumatoid arthritis based on cyclic citrullinated peptide. *J. Pharm. Biomed. Anal.* **115**, 107–113 (2015).
68. Ru, Z. et al. A new pattern of citrullinated peptides improves the sensitivity for diagnosing rheumatoid arthritis. *Clin. Biochem.* **105–106**, 87–93 (2022).
69. Roman-Melendez, G. D. et al. Citrullination of a phage-displayed human peptidome library reveals the fine specificities of rheumatoid arthritis-associated autoantibodies. *EBioMedicine* **71**, 103506 (2021).
70. Huang, D. & Kraus, W. L. The expanding universe of PARP1-mediated molecular and therapeutic mechanisms. *Mol. Cell* **82**, 2315–2334 (2022).
71. Lewis, H. D. & Nacht, M. IPAD or PADI—‘tablets’ with therapeutic disease potential? *Curr. Opin. Chem. Biol.* **33**, 169–178 (2016).
72. Lange, S. et al. Peptidylarginine deiminases—roles in cancer and neurodegeneration and possible avenues for therapeutic intervention via modulation of exosome and microvesicle (EMV) release? *Int. J. Mol. Sci.* **18**, 1196 (2017).
73. Gudmann, N. S., Hansen, N. U. B., Jensen, A. C. B., Karsdal, M. A. & Siebuhr, A. S. Biological relevance of citrullinations: diagnostic, prognostic and therapeutic options. *Autoimmunity* **48**, 73–79 (2015).
74. Matsuo, K. et al. Identification of novel citrullinated autoantigens of synovium in rheumatoid arthritis using a proteomic approach. *Arthritis Res. Ther.* **8**, R175(2006).
75. Li, P. et al. PAD4 is essential for antibacterial innate immunity mediated by neutrophil extracellular traps. *J. Exp. Med.* **207**, 1853–1862 (2010).
76. Harding, C. R. & Scott, I. R. Histidine-rich proteins (filaggrins): structural and functional heterogeneity during epidermal differentiation. *J. Mol. Biol.* **170**, 651–673 (1983).
77. Méchin, M. C. et al. Update on peptidylarginine deiminases and deimination in skin physiology and severe human diseases. *Int. J. Cosmet. Sci.* **29**, 147–168 (2007).
78. Bernstein, B. E. et al. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169–181 (2005).
79. Zhang, X., Gamble, M. J., Stadler, S., Cherrington, B. D. & Causey, C. P. Genome-wide analysis reveals PADI4 cooperates with Elk-1 to activate c-Fos expression in breast cancer cells. *PLoS Genet.* **7**, 1002112 (2011).
80. Ochoa, D. et al. The functional landscape of the human phosphoproteome. *Nat. Biotechnol.* **38**, 365–373 (2020).

81. Schölz, C. et al. Acetylation site specificities of lysine deacetylase inhibitors in human cells. *Nat. Biotechnol.* **33**, 415–425 (2015).
82. Darrach, E. & Andrade, F. Rheumatoid arthritis and citrullination. *Curr. Opin. Rheumatol.* **30**, 72–78 (2018).
83. Fox, D. A. Citrullination: a specific target for the autoimmune response in rheumatoid arthritis. *J. Immunol.* **195**, 5–7 (2015).
84. Bettner, L. F. et al. Combinations of anticyclic citrullinated protein antibody, rheumatoid factor, and serum calprotectin positivity are associated with the diagnosis of rheumatoid arthritis within 3 years. *ACR Open Rheumatol.* **3**, 684–689 (2021).
85. Bukhari, M. et al. The performance of anti-cyclic citrullinated peptide antibodies in predicting the severity of radiologic damage in inflammatory polyarthritis: results from the Norfolk Arthritis Register. *Arthritis Rheum.* **56**, 2929–2935 (2007).
86. Madan, P. S., Savur, A. D., Kamath, S. U., Baliga, S. & Mallya, S. A study of correlation of disease severity with antibody titers in rheumatoid arthritis. *J. Clin. Diagn. Res.* **11**, OC09–OC13 (2019).
87. Smolen, J. S., Aletaha, D. & McInnes, I. B. Rheumatoid arthritis. *Lancet* **388**, 2023–2038 (2016).
88. Zhao, X. et al. Circulating immune complexes contain citrullinated fibrinogen in rheumatoid arthritis. *Arthritis Res. Ther.* **10**, R94–R94 (2008).
89. Tanikawa, C. et al. Citrullination of RGG motifs in FET proteins by PAD4 regulates protein aggregation and ALS susceptibility. *Cell Rep.* **22**, 1473–1483 (2018).
90. Meyer, J. G. et al. Quantification of lysine acetylation and succinylation stoichiometry in proteins using mass spectrometric data-independent acquisitions (SWATH). *J. Am. Soc. Mass. Spectrom.* **27**, 1758–1771 (2016).
91. Wu, R. et al. A large-scale method to measure absolute protein phosphorylation stoichiometries. *Nat. Methods* **8**, 677–683 (2011).
119. Sun, S. & Zhang, H. Large-scale measurement of absolute protein glycosylation stoichiometry. *Anal. Chem.* **87**, 6479–6482 (2015).
120. Lewallen, D. M. et al. Chemical proteomic platform to identify citrullinated proteins. *ACS Chem. Biol.* **10**, 2520–2528 (2015).
121. Chaerkady, R. et al. Characterization of citrullination sites in neutrophils and mast cells activated by ionomycin via integration of mass spectrometry and machine learning. *J. Proteome Res.* **20**, 3150–3164 (2021).

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

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

## Methods

Our research complies with all relevant ethical regulations. The samples used in this study were obtained following informed consent from individuals recruited at the Center for Rheumatology and Spine Diseases, Copenhagen University Hospital Glostrup, after approval by the local ethical committee (approval ID H-16042831). The local ethics committee was: De Videnskabetiske Komiteer for Region Hovedstaden, Regionsgården Kongens Vænge 2, 3400 Hillerød, Denmark.

### Cell culture

HL60 cells were grown in RPMI medium (cat. no. 21875091, Gibco) supplemented with 10% fetal bovine serum (FBS) and penicillin–streptomycin (100 U ml<sup>-1</sup>) (Gibco) at 37 °C and 5% CO<sub>2</sub>. The HL60 cell line was a gift from M. Christophorou. A proportion of the cells were differentiated over the course of 4 d to neutrophil-like cells by the addition of 1.25% (vol/vol) DMSO to the medium. The experiments were performed in biological (cell culture) triplicate. Cells were washed in 37 °C PBS and transferred to 37 °C Locke's solution (10 mM Hepes pH 7.5, 150 mM NaCl, 5 mM KCl and 2 mM CaCl<sub>2</sub>, 0.1% glucose) at a concentration of 2 × 10<sup>6</sup> cells ml<sup>-1</sup>. Citrullination of proteins was induced via activation of PADI4 by the addition of calcium ionophore Af23187 (cat. C7522, Merck) to a final concentration of 4 μM, for either 15 or 30 min, at 37 °C. Control samples, of both undifferentiated HL60 cells and differentiated neutrophil-like cells, were collected from Locke's solution before calcium treatment for further sample preparation.

### PADI4 inhibition by GSK484

Quadruplicate cultures of NLCs were incubated in Locke's solution, as described above, with a range of GSK484 (cat. no. SML1658, Merck) concentrations at 37 °C for 30 min before calcium activation, using 4 μM calcium ionophore A23187 (cat. no. C7522, Merck). The GSK484 concentrations were 1 μM, 5 μM and 20 μM, in addition to a DMSO control.

### Cell lysis and protein digestion

The cell pellets were lysed in lysis buffer (6 M guanidine-HCl, 50 mM TRIS, pH 8.5) and further processed using standard sample-preparation methods. The proteins were digested by two rounds of Lys-C digestion (see Supplementary Note 6 for details).

### Purification of peptides

Peptides were purified using reversed-phase C18 cartridges (SepPak Classic, 350 mg, Waters). Cartridges were activated with 5 ml acetonitrile (ACN) and equilibrated three times with 5 ml of 0.1% TFA, after which samples were loaded. Sample loading was accelerated using a vacuum manifold, maintaining two-thirds atmospheric pressure. Following loading, cartridges were washed three times with 5 ml of 0.1% TFA, after which peptides were eluted using 4 ml of 30% ACN in 0.1% TFA. The eluted peptides were frozen overnight at –80 °C in 15-ml tubes with small holes punctured into the caps, after which the frozen peptides were lyophilized for 96 h. Lyophilized peptides were dissolved in 25 mM ammonium bicarbonate pH 8.5, and the peptide concentration was estimated through absorbance at 280 nm, using a NanoDrop instrument.

### Offline high-pH reversed-phase HPLC fractionation

For each experimental replicate, 0.6 mg peptide was fractionated into 46 fractions using an XBridge BEH130 C18 3.5 μm 4.6 mm × 250 mm column (Waters) on an Ultimate 3000 HPLC system (Dionex), operating at a flow rate of 1 ml min<sup>-1</sup>. The flow was composed of three buffers: buffer A (Milli-Q water), buffer B (100% ACN) and buffer C (25 mM ammonium hydroxide). Prior to loading, samples were basified to pH > 10 by the addition of ammonium hydroxide (gradient details are available in Supplementary Note 7). Collected fractions were transferred to Eppendorf Protein LoBind tubes with small holes punctured

in the caps, and frozen at –80 °C overnight. The frozen samples were lyophilized for 96 h and afterwards were dissolved in 1% formic acid (FA).

### Mass spectrometry analysis

Samples were measured using a Q-Exactive HF-X mass spectrometer or an Exploris 480 mass spectrometer (Thermo Fisher Scientific). Peptides were separated by online reversed-phase liquid chromatography using an EASY-nLC 1200 system (Thermo Fisher Scientific), using a 15-cm-long analytical column with an internal diameter of 75 μm, packed in-house using ReproSil-Pur 120 C18-AQ 1.9 μm beads (Dr. Maisch). The analytical column was heated to 40 °C using a column oven, and peptides were eluted from the column using a gradient of buffer A (0.1% FA) and buffer B (80% ACN in 0.1% FA). The gradient ranged from 4% to 38% buffer B over 30 min, followed by an increase to 90% buffer B over 4 min to ensure elution of all peptides, followed by a washing block of 6 min.

For the pilot experiment, performed on the Q-Exactive HF-X instrument, electrospray ionization was achieved using a Nanospray Flex Ion Source (Thermo Fisher Scientific). The spray voltage was set to 2 kV, the capillary temperature to 275 °C and the radio frequency level to 40%. Full scans were performed at a resolution of ×60,000, with a scan range of 300 to 1,750 *m/z*, a maximum injection time of 60 ms and an automatic gain control (AGC) target of 3,000,000 charges. Precursors were isolated at a width of 1.3 *m/z*, with an AGC target of 200,000 charges. Repeated sequencing of selected precursors was excluded by dynamic exclusion of 60 s. Precursor fragmentation was achieved using higher energy collision dissociation (HCD). MS/MS data were measured using the Orbitrap with a maximum injection time of 90 ms and a resolution of 4 × 5,000. The Top9 data-dependent MS/MS method was used to acquire MS data.

For the optimized GSK experiment, performed on the Exploris 480 instrument, electrospray ionization was achieved using a NG Ion Source (Thermo Fisher Scientific). The spray voltage was set to 2 kV, the capillary temperature to 275 °C and the radio frequency level to 40%. Full scans were performed at a resolution of ×120,000, with a scan range of 300 to 1,750 *m/z*, the maximum injection time set to 'Auto' and the normalized AGC target set to '200' (2,000,000 charges). Precursors were isolated at a width of 1.3 *m/z*, with a normalized AGC target of '200' (200,000 charges). Repeated sequencing of selected precursors was excluded by dynamic exclusion of 60 s. Precursor fragmentation was achieved using HCD. MS/MS data were measured using the Orbitrap with a maximum injection time set to 'Auto' and a resolution of ×30,000.

### Western blotting

Cell pellets were lysed in SDS Lysis Buffer (2% SDS, 50 mM Tris-HCl pH 8.5, 150 mM NaCl) and homogenized by heating to 99 °C and shaking at 1,400 r.p.m. for 30 min. Protein concentrations across lysates were equalized using the Pierce BCA Protein Assay Kit (cat. no. 23225, Pierce), according to the manufacturer's instructions. An immunoblot was performed using standard approaches with an Invitrogen chamber and blot module. Proteins were transferred to a PVDF membrane (Immobilon) for 90 min at 0.4 A. Membranes were blocked using 5% milk (Fluka Analytical) in PBS supplemented with Tween-20 (0.1%; PBST) or 5% bovine serum albumin (BSA) in PBST, following the antibody manufacturer's recommendations. The following antibodies were used: rabbit polyclonal PADI4 (1:1,000, cat. no. P4749, Sigma Aldrich), rabbit monoclonal H3 (citrulline Arg2) antibody (1:1,000, cat. no. Ab176843, clone EPR17703, Abcam) and rabbit polyclonal GAPDH (1:1,000, Ab9485, Abcam). The Anti-Citrulline (Modified) Detection Kit (cat. no. 17-347B, Merck) with the anti-modified citrulline antibody (part number MABS487, clone C4) was used to measure global citrullination<sup>24</sup>, according to the manufacturer's instructions.

### Mass spectrometry data analysis

The raw mass spectrometry data files were analyzed using MaxQuant software (version 1.5.3.30), a freely available software routinely used in this field. MaxQuant settings used for analysis are available in Supplementary Note 8. A HUMAN.fasta database was extracted from UniProt on 5 May 2020 to serve as a theoretical spectral library. The HUMAN.fasta database contained 96,821 protein entries.

### Mass spectrometry data filtering

In addition to automatic filtering and FDR control, as applied by MaxQuant, the data were manually filtered to ensure proper identification and localization of citrullination. Citrullination-site identification was allowed only if the localization probability was >0.90. For quantification of citrullination, further PSMs were accepted with a localization of >0.75, as this is the standard cut-off in proteomics; the 0.90 used in this study for qualification is stringent. MaxQuant intensity values (a quantitative metric corresponding to peak area-under-the-curve at the MS1 level) were manually transferred from the evidence file (by mapping on experiment and fixed evidence IDs) to the citrullination sites using only PSMs with a localization of >0.75.

LFQ intensities were normalized within each condition and missing values were imputed across replicates using Perseus software.

### Mass spectrometry statistical analysis

Statistical handling of the data and hierarchical clustering was primarily performed using the freely available Perseus software (version 1.6.14.0)<sup>92</sup>. Significantly enriched Gene Ontology terms were determined using the Functional Annotation Tool of the DAVID Bioinformatics database<sup>93,94</sup>. Venn diagrams were generated using the online DeepVenn program<sup>95</sup>. Boxplots were generated using the BoxPlotR web tool<sup>96</sup>. Kinase–substrate relationships were predicted using the online NetworKIN tool<sup>97,98</sup>. The sequence motif was generated using the iceLogo software (version 1.2)<sup>99</sup>, with background sequences extracted from non-citrullinated arginine residues in all citrullinated proteins.

### Analysis of transcription factor target genes and gene-set enrichment

For each citrullinated and non-citrullinated transcription factor (TF), all target genes were retrieved from TFEA.ChIP using the ReMap 2020 (ref. 100) and GeneHancer Double Elite<sup>101</sup> data set. Data were available for 115 citrullinated TFs and 233 non-citrullinated TFs. For each of the 16,544 target genes with an Ensembl annotation, the fraction of citrullinated and non-citrullinated TFs that regulate it was determined (Supplementary Table 6). The final score was calculated as the logarithm of the ratio of citrullinated versus non-citrullinated fractions. This value is negative for a given target if there are more non-citrullinated TFs regulating the target than citrullinated ones, and vice versa. All targets were ordered by this value and the whole ranked list was used as input for the STRING v11 gene-set enrichment analysis<sup>102</sup>. The resulting enriched annotations with a FDR below 0.05 are provided in Supplementary Table 7. A given annotation describes the citrullinated TFs if it is enriched at the bottom of the input, and the non-citrullinated TFs if it is enriched at the top of the input.

### Enrichment analysis of regions targeted for citrullination

For the analysis of citrullination sites in disordered regions, disorder was predicted using IUPred2A<sup>103</sup> for all identified sequences. The set of long disordered regions was obtained using IUPred2A's long disorder option and a minimum region length of 31 consecutive residues with a prediction score of ≥0.5. Regions predicted using the short disorder option were retained if they contained 2 to 30 consecutive residues with a score of ≥0.5. Predictions from all identified sequences were included in the analysis of disordered regions.

Annotations of domains and NLSs were obtained in gff format from UniProtKB<sup>104</sup>, and their sequences were derived from the sequences

identified in the MS run. All identified sequences with any kind of feature annotation were included in the domain and NLS analyses.

For each feature category (long and short disorder, domain, NLS), a Fisher's exact test was performed, using the counts of unmodified arginine residues and citrullinated arginine residues inside and outside the features. Fold enrichment of citrullines inside features compared with citrullines outside features was calculated as:

$$FE = \frac{Cit_i R_o}{R_i Cit_o},$$

where  $Cit_i$  and  $Cit_o$  refer to the numbers of citrullines inside and outside features, respectively, and  $R_i$  and  $R_o$  refer to the total number of unmodified and modified arginine residues inside and outside features.

### Peptide microarray screen using clinical rheumatoid arthritis samples

Peptide microarrays were produced in collaboration with PEPperPRINT, according to their custom PEPperCHIP Discovery microarrays workflow. Briefly, linear peptides were synthesized step-wise onto a coated glass slide using a precision laser, allowing the incorporation of non-canonical amino acids, such as citrulline. The peptide arrays each contained 75,460 17-amino-acid peptides from the following groups: citrullinated sites and their unmodified counterparts, peptides with multiple citrullination sites on the same peptide and their unmodified counterparts, randomly selected arginine residues from the same protein groups and their citrullinated counterparts, known ACPA reactive peptide sequences and their unmodified counterparts<sup>20,67,105–115</sup> and a commercially available PEPperPRINT CCP array (see Supplementary Tables 8–15). Owing to the number of controls included, the citrulline sites were split across two chips containing the top 50% sites and bottom 50% sites, respectively, as determined by intensities derived from our MS experiment.

The peptide microarrays were incubated with synovial fluids pooled from six individuals for each of the three following groups: anti-CCP-positive RA; anti-CCP-negative RA; and ankylosing spondylitis. All participants were recruited at Center for Rheumatology and Spine Diseases, Copenhagen University Hospital Glostrup, after informed consent was obtained and under approval by the local ethical committee (H-16042831). Of the 18 participants, 12 were females and 6 were males. Participant age at the point of sample collection ranged from 28 to 76 years old. Unspecific binding was tested by incubating the microarray alone with the secondary antibody used for primary antibody detection. Before incubation, arrays were blocked with Rockland blocking buffer MB-070, after which the synovial fluid was diluted in the buffer and was allowed to bind to the chip for 16 h at 4 °C under mild shaking at 140 r.p.m. After incubation, the arrays were washed three times with PBS supplemented with 0.05% Tween-20 after each incubation step.

To detect binding of clinical anti-citrulline protein antibodies, the intensity of the secondary antibody goat anti-human-IgG (Fc) DyLight680 (0.1 μg ml<sup>-1</sup>) was recorded using the LI-COR Odyssey Imaging System with the following parameters: scanning offset 0.67 mm, resolution 21 μm, scanning intensities of 8/8 (red 680 nm; green 800 nm).

To quantify the intensity of the resulting spots, 16-bit grayscale tiff files of the arrays were recorded and analyzed the images with PepSlide Analyzer (Sicasys software).

### Peptide microarray data analysis

To analyze the peptide microarray data, we used the raw data provided by PEPperPRINT ('Raw Data' tabs in Supplementary Tables 8–15). The 'Raw Mean' intensity was used for further processing, as this represents the most basic readout and prevents any 'O' reads, which would have to be imputed at a later stage. Within each array, the median was calculated for all reads corresponding to the same peptide sequences,

which in most cases was for  $n = 2$  (duplicate prints), but in some cases was  $n = 4, 6, 8$  or  $10$  for peptides in multiple classes that were printed several times on the same chip. All values were  $\log_2$ -transformed, at which point the global median of the pairs of chips treated with the same synovial fluid were normalized to each other. Next, we calculated ratios of change for all pairs of citrullinated peptides versus their unmodified counterpart peptide. At this stage, we computed the median for the ratios of peptide pairs that existed on both chips for the same synovial fluid mixture. This process was applied to only the controls, because they were printed in technical duplicates on two chips (as well as technical duplicates within each chip, as outlined above). This resulted in four data points for each citrulline-baseline pair, stratified only by synovial fluid type: second antibody control, disease control, anti-CCP negative, and anti-CCP positive. Finally, we calculated the  $z$ -score for each peptide pair. Supplementary Note 9 contains details on the statistical analysis of microarray data.

IceLogo v1.3.8 was used to generate sequence logos. As a foreground, we used the top 1,000 citrullinated arginine residues, identified by MS, overlapping either with the most enriched arginine residues in the anti-CCP-positive group (as derived from the microarray data) or the most enriched arginine residues in the anti-CCP-negative group. As a background, the other ~13,000 MS-identified citrullinated arginine residues were used. The  $P$  value for iceLogo analysis was set at 5% (default), and the s.d. was used to visualize the magnitude of change.

### NetSurfP-3.0 structural predictions

The NetSurfP3.0 model<sup>164</sup> was used to predict the RSA for all residues in proteins identified as citrullinated in this study. To this end, the sequences for all citrullinated proteins were submitted in small batches to the 'NetSurfP - 3.0' web service at the Technical University of Denmark. All batches were collated, and the predicted values for arginine residues were extracted. For statistical comparison of the data, all foreground arginine residues (that is those found to be citrullinated) were compared versus all background arginine residues (that is, all non-modified arginine residues from the same proteins). RSA values ranged between 0 and 1, and were grouped into bins of 0.02 for visualization. Significance was tested using two-tailed Spearman nonparametric correlation analysis.

### Prediction of accessibility of citrullination sites

The accessibility of the citrullination sites was estimated from the AlphaFold<sup>116</sup>, predicted human protein structure database. For each of the AlphaFold structures, the exposure of all arginine residues was found. The exposure was estimated as the coordination number (CN). The CN is the number of  $\alpha$  atoms within a sphere around the  $\alpha$  atom, and for this study, a radius of 13 Å was applied. The CN was calculated using the Python program hsexpo from the Bio.PDB library<sup>117</sup> and collected in a data set. For statistical comparison of the data, all foreground arginine residues (that is, those found to be citrullinated) were compared with all background arginine residues (that is, all non-modified arginine residues from the same proteins). Significance was tested using two-tailed Spearman nonparametric correlation analysis.

### Structural analysis via AlphaFold

All published AlphaFold predictions (.cif files) for human proteins were downloaded<sup>116</sup>, the files were parsed to output a text file containing the protein Uniprot ID, amino acid residue, position in sequence, pLDDT score, F number and assigned structure for every residue in the library, and pLDDT values (model confidence) and secondary structure elements were parsed out for all arginine residues. If the AlphaFold predictions consisted of overlapping segments (as is the case for large proteins), the median pLDDT was taken for any residues predicted across multiple segments. For secondary structure elements, if there was ambiguity, the most confidently predicted elements (going by pLDDT) were taken. Next, we defined two background groups, all

arginines in human proteins, and only arginines in proteins mapped by MS to be citrullinated in this study. Out of all MS-identified sites, 13,633 sites could be aligned to the AlphaFold predictions, that is pLDDT and secondary structure elements could be derived for these 13,633 citrullinated arginine residues. The missing sites were either not predicted by AlphaFold, had a mismatching protein ID (because of UniProt database changes over time), were exclusive to an isoform (AlphaFold predicted only canonical entries) or did not match to an arginine (because of UniProt database changes over time). pLDDT values were visualized as boxplots for several selections of arginine residues, using BoxPlotR<sup>96</sup> for visualization. For investigation of enrichment of specific secondary structure elements, the Fisher exact test was performed to compare groups of secondary structure elements to groups of citrullination sites, using either all arginines in human proteins or all arginines in citrullinated proteins as a background. Correction for multiple-hypotheses testing was performed using Benjamini–Hochberg adjustment of the  $P$  values.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium through the PRIDE<sup>118</sup> partner repository with the data set identifier PXD038702. Statistical source data for all main figures and Extended Data figures can be found in the provided source data files in addition to Extended Data tables. Databases used in no particular order: AlphaFold<sup>116</sup>, UniProtKB<sup>104</sup>, DAVID Bioinformatics<sup>94</sup>, AAgAtlas 1.0 (ref. 45), ReMap2020 (ref. 100) and GeneHancer Double Elite<sup>101</sup>. Source data are provided with this paper.

### References

- Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).
- Sherman, B. T. et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* **50**, W216–W221 (2022).
- Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
- Hulsen, T., de Vlieg, J. & Alkema, W. BioVenn—a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* **9**, 488 (2008).
- Spitzer, M., Wildenhain, J., Rappsilber, J. & Tyers, M. BoxPlotR: a web tool for generation of box plots. *Nat. Methods* **11**, 121–122 (2014).
- Horn, H. et al. KinomeXplorer: an integrated platform for kinome biology studies. *Nat. Methods* **11**, 603–604 (2014).
- Linding, R. et al. NetworkKIN: a resource for exploring cellular phosphorylation networks. *Nucleic Acids Res.* **36**, D695 (2008).
- Ham, A. J. L. et al. Improved visualization of protein consensus sequences by iceLogo. *Nat. Methods* **6**, 786–787 (2009).
- Cheneby, J. et al. ReMap 2020: a database of regulatory regions from an integrative analysis of human and *Arabidopsis* DNA-binding sequencing experiments. *Nucleic Acids Res.* **48**, D180–D188 (2020).
- Fishilevich, S. et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database* **2017**, bax028 (2017).
- Szklarczyk, D. et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).

103. Mészáros, B., Erdős, G. & Dosztányi, Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **46**, W329–W337 (2018).
104. Bateman, A. et al. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).
105. Hansson, M. et al. Validation of a multiplex chip-based assay for the detection of autoantibodies against citrullinated peptides. *Arthritis Res. Ther.* **14**, R201 (2012).
106. Bacic, L. et al. Structure and dynamics of the chromatin remodeler ALC1 bound to a PARylated nucleosome. *eLife* **10**, e71420 (2021).
107. Wang, F. et al. Identification of citrullinated peptides in the synovial fluid of patients with rheumatoid arthritis using LC-MALDI-TOF/TOF. *Clin. Rheumatol.* **35**, 2185–2194 (2016).
108. Bennike, T. et al. Optimizing the identification of citrullinated peptides by mass spectrometry: utilizing the inability of trypsin to cleave after citrullinated amino acids. *J. Proteom. Bioinform.* **612**, 288–295 (2013).
109. Wood, D. D. & Moscareilos, M. A. The isolation, characterization, and lipid-aggregating properties of a citrulline containing myelin basic protein. *J. Biol. Chem.* **264**, 5121–5127 (1989).
110. Almaguel, F. A., Sanchez, T. W., Ortiz-Hernandez, G. L. & Casiano, C. A. Alpha-enolase: emerging tumor-associated antigen, cancer biomarker, and oncotherapeutic target. *Front. Genet.* **11**, 614726 (2021).
111. Vossenaar, E. R. & van Venrooij, W. J. Citrullinated proteins: sparks that may ignite the fire in rheumatoid arthritis. *Arthritis Res. Ther.* **6**, 107–111 (2004).
112. Verpoort, K. N. et al. Fine specificity of the anti-citrullinated protein antibody response is influenced by the shared epitope alleles. *Arthritis Rheum.* **56**, 3949–3952 (2007).
113. Jang, B. et al. Vimentin citrullination probed by a novel monoclonal antibody serves as a specific indicator for reactive astrocytes in neurodegeneration. *Neuropathol. Appl. Neurobiol.* **46**, 751–769 (2020).
114. Bang, H. et al. Mutation and citrullination modifies vimentin to a novel autoantigen for rheumatoid arthritis. *Arthritis Rheum.* **56**, 2503–2511 (2007).
115. Darrah, E. et al. Erosive rheumatoid arthritis is associated with antibodies that activate PAD4 by increasing calcium sensitivity. *Sci. Transl. Med.* **5**, 186ra65 (2013).
116. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
117. Hamelryck, T. & Manderick, B. PDB file parser and structure class implemented in Python. *Bioinformatics* **19**, 2308–2310 (2003).
118. Perez-Riverol, Y. et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **50**, D543–D552 (2022).

## Acknowledgements

The work carried out in this study was in part supported by the Novo Nordisk Foundation Center for Protein Research, the Novo

Nordisk Foundation (grant agreement numbers NNF14CC0001 and NNF13OC0006477), The Danish Council of Independent Research (grant agreement numbers 4002-00051, 4183-00322A and 8020-00220B) and The Danish Cancer Society (grant agreement R146-A9159-16-S2). The proteomics technology was part of a project that has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement EPIC-XS-823839. We thank members of the NNF-CPR Mass Spectrometry Platform for instrument support and technical assistance. M.A.C. was supported by a Sir Henry Dale Fellowship, jointly funded by the Wellcome Trust and the Royal Society (grant no. 105642/A/14/Z) and by funding from the UK Biotechnology and Biological Sciences Research Council (BBS/E/B/000C0421). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the paper.

## Author contributions

A.S.R. and M.L.N. designed the experiment. M.A.C. provided input on the project, experimental samples and training on cell handling. A.S.R. prepared the immunoblots. A.S.R. and S.C.B.-L. prepared all MS experiments, and A.S.R., S.C.B.-L. and I.A.H. measured all samples on the mass spectrometer and optimized the MS workflow. A.S.R., I.A.H., J.D.E., C.L. and M.R. performed bioinformatics and statistical analyses. N.T.D. and L.J.J. performed transcription factor analysis. I.A.H., R.K. and L.J.J. performed structural predictions. I.A.H. and J.D.E. designed and analyzed the peptide microarray experiments in collaboration with PEPperPRINT. C.H.N., L.T. and D.D. collected the synovial fluid samples from people with RA and control participants. M.L.N. supervised the project. A.S.R. and M.L.N. wrote the paper with input from all authors.

## Competing interests

The authors declare no competing interests.

## Additional information

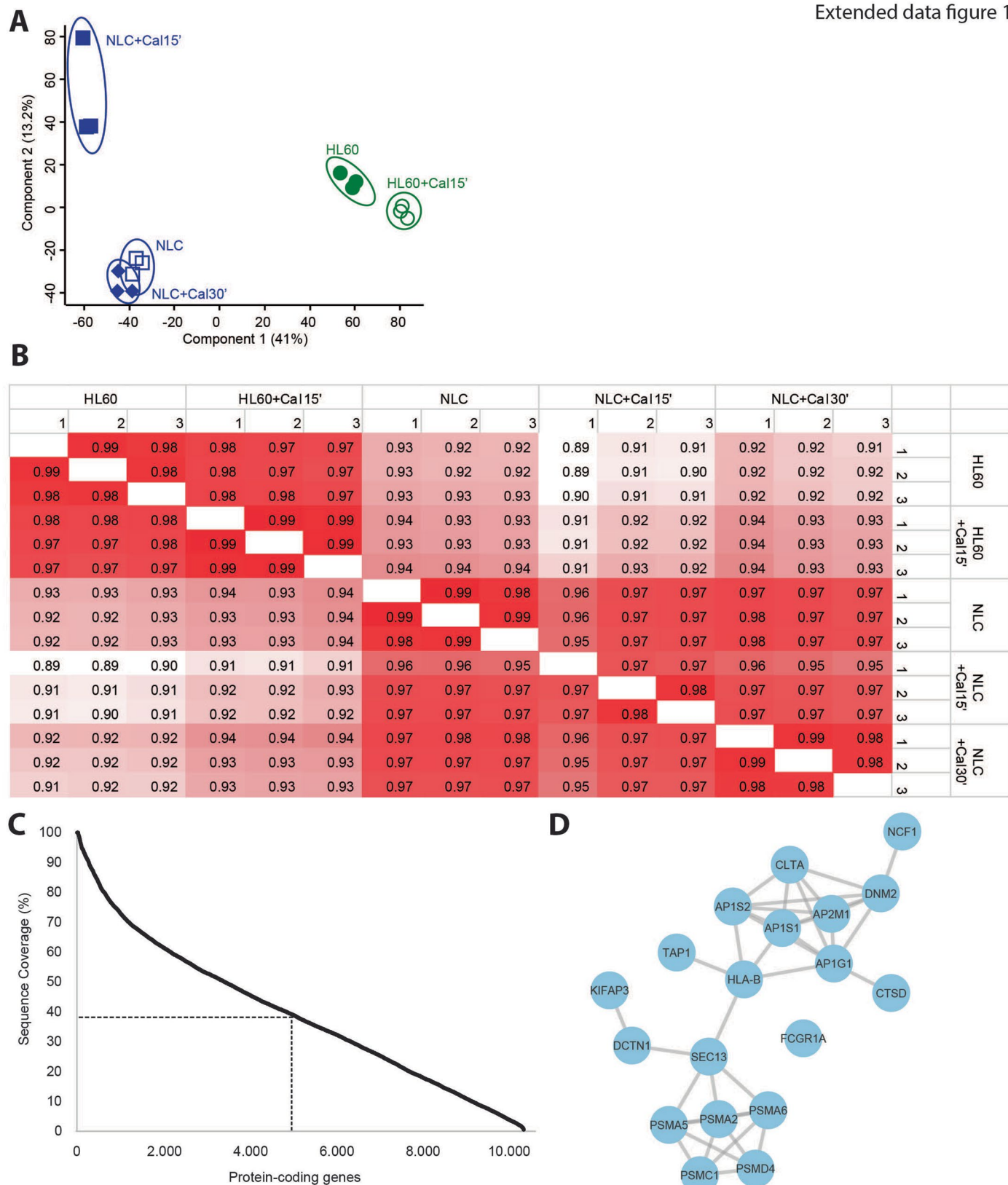
**Extended data** is available for this paper at <https://doi.org/10.1038/s41594-024-01214-9>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41594-024-01214-9>.

**Correspondence and requests for materials** should be addressed to Michael L. Nielsen.

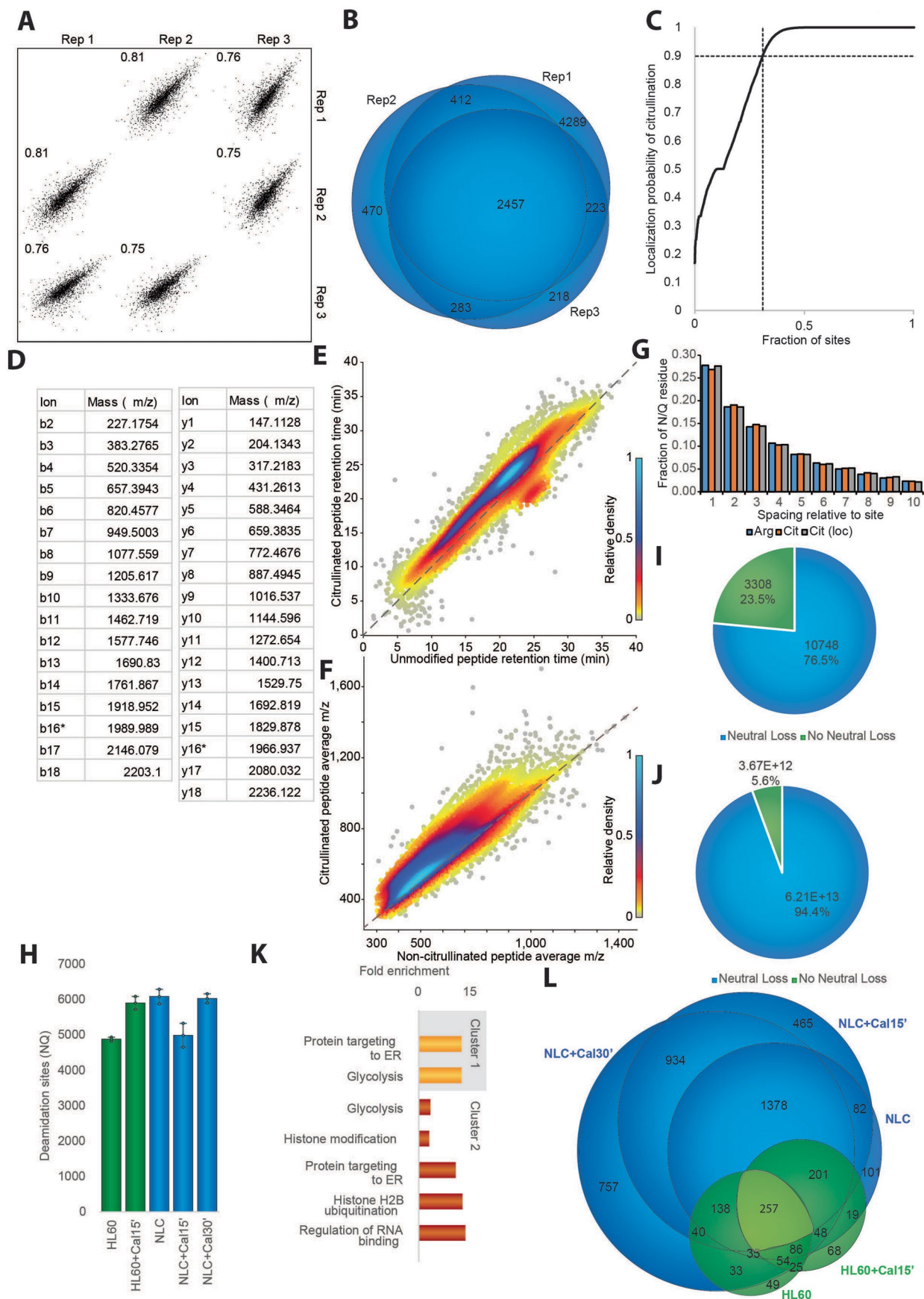
**Peer review information** *Nature Structural & Molecular Biology* thanks Jeroen Krijgsveld and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Dimitris Typas, in collaboration with the *Nature Structural & Molecular Biology* team.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



**Extended Data Fig. 1 | Validation of experimental system.** **a**) Principal component analysis of all conditions at protein level. Eigenvalues are displayed on the axes. **b**) Pearson correlation of proteins detected in all replicates of experimental conditions. Clear distinction detection in proteome following DMSO-differentiation. **c**) Sequence coverage of proteins detected. Dotted

line indicates median. **d**) STRING network visualizing functional associations between proteins annotated to be involved in antigen processing and presentation detected as upregulated in NLCs (Cluster 1, Fig. 1i, k). Default STRING confidence score cutoff of > 0.4 was used.

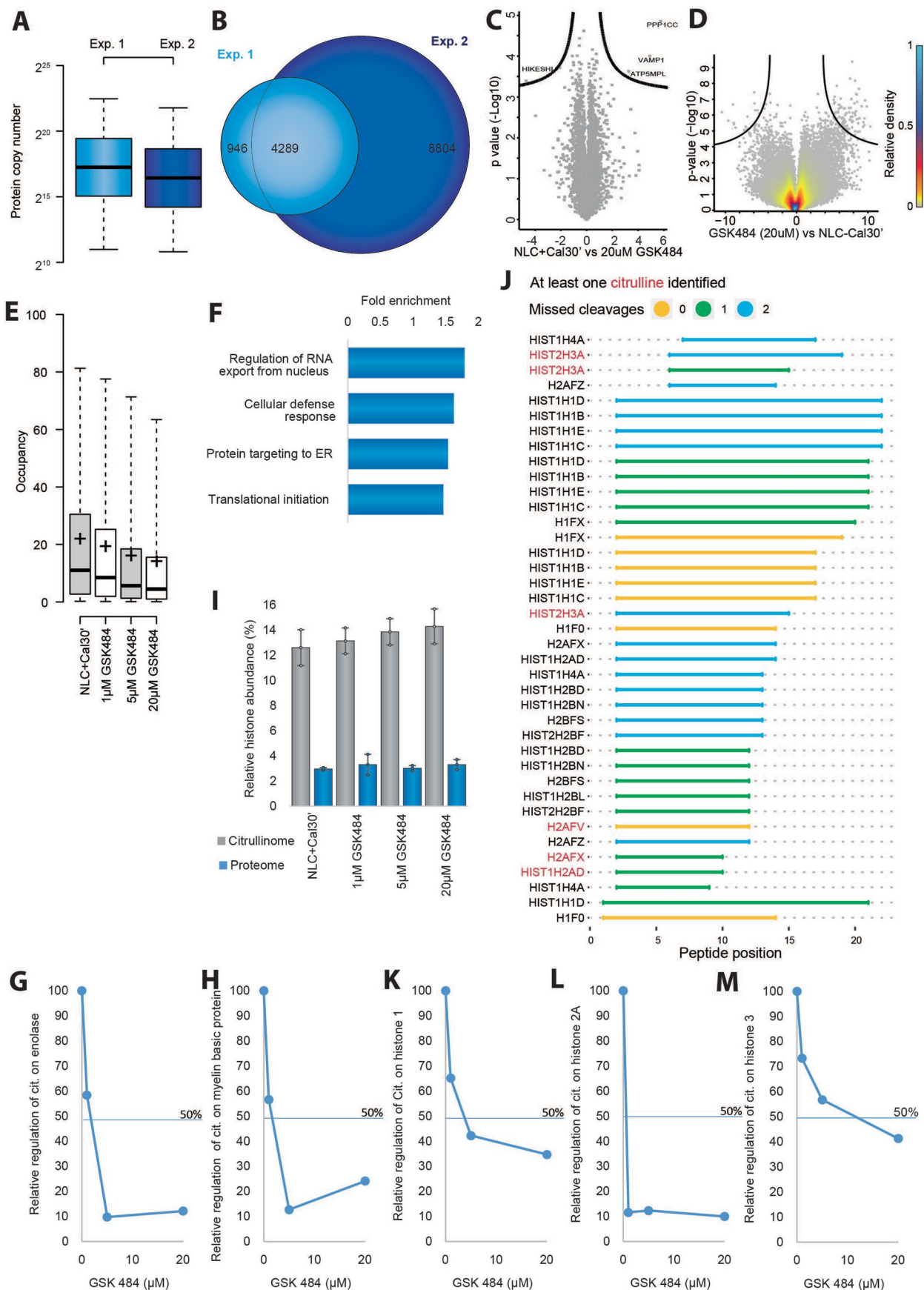


Extended Data Fig. 2 | See next page for caption.

**Extended Data Fig. 2 | Validation of detection of citrullination sites.**

**a)** Pearson correlation of citrullination sites detected in all three replicates of NLC+Cal30' condition. **b)** Venn diagram showing site detection overlap between all three replicates of NLC+Cal30' condition. **c)** Citrullination localization probability plotted against the ranked fraction of all peptide-spectrum-matches (PSMs). Despite all probabilities being displayed, only those above 0.9 were used for assignment of unique citrullinated peptides and sites. **d)** Mass of b and y ions shown in MS/MS spectrum in Fig. 2c. **e)** Density plot of elution time during low pH reversed-phase separation for unmodified peptides (x-axis) vs citrullinated peptides (y-axis). **f)** Density plot of detected m/z for unmodified peptides (x-axis) vs m/z for citrullinated peptides (y-axis). **g)** Fraction of N/Q residues relative to

citrullination site spacing for unmodified arginine residues, citrullination sites and localized citrullination sites. **h)** Number of deamidation sites of asparagine (N) and glutamine (Q) detected across conditions. n = 3 biological independent samples. Data are presented as mean values  $\pm$  SD. **i)** Pie chart showing distribution of sites detected with neutral loss of cyanic acid or no neutral loss. **j)** Pie chart showing citrullination abundance associated with sites for which a neutral loss was either detected or not detected. **k)** Gene Ontology (GO) term enrichment analysis for cellular components of citrullination target proteins in selections displayed in 1 G (cluster 1 and 2), as compared to the human proteome. Statistical information in Supplementary Table 5. **l)** Scaled venn diagram displaying overall of number sites detected in each experimental condition.

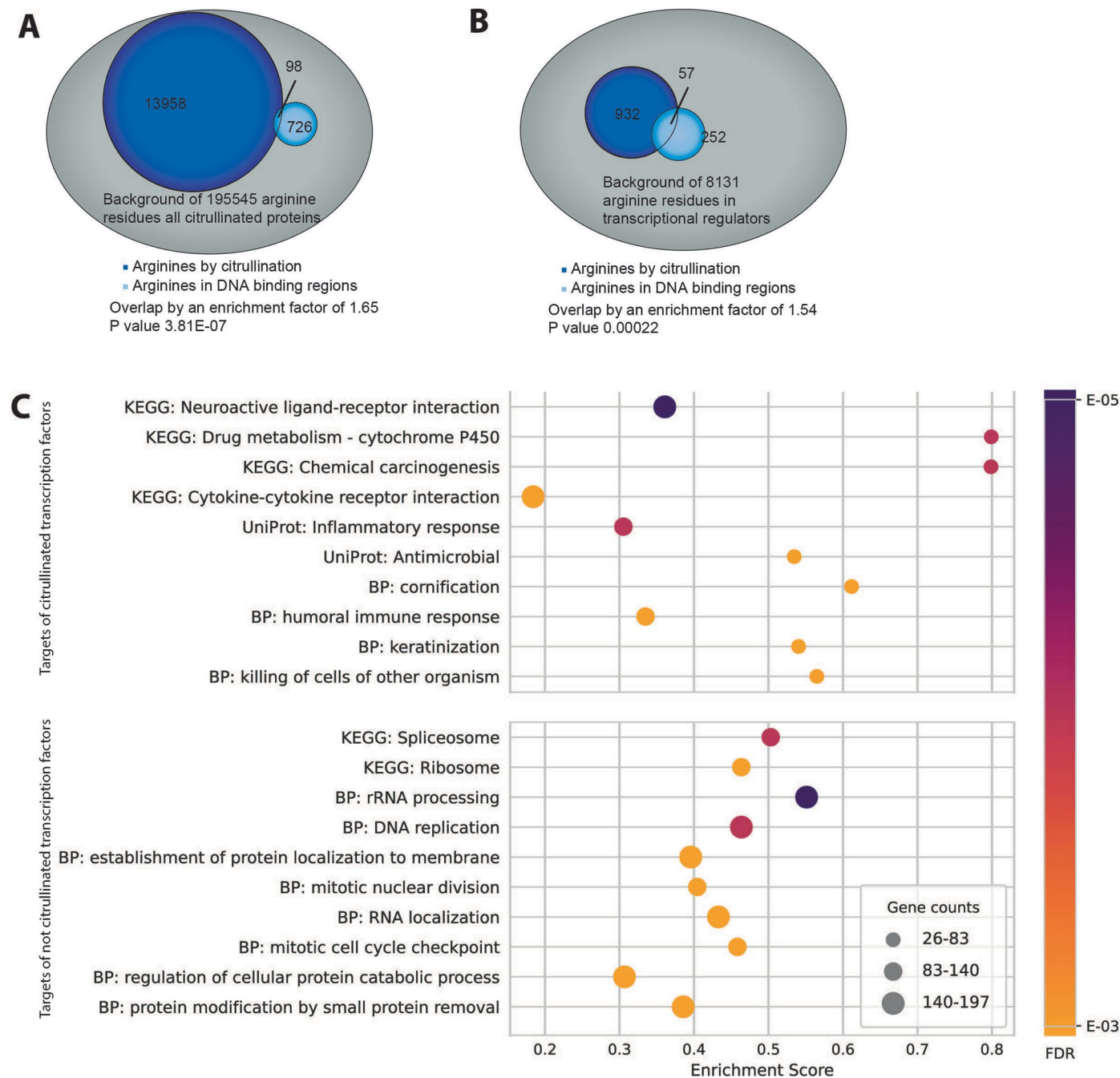


Extended Data Fig. 3 | See next page for caption.

**Extended Data Fig. 3 | Regulation of PADI4 by specific inhibitor, GSK484.**

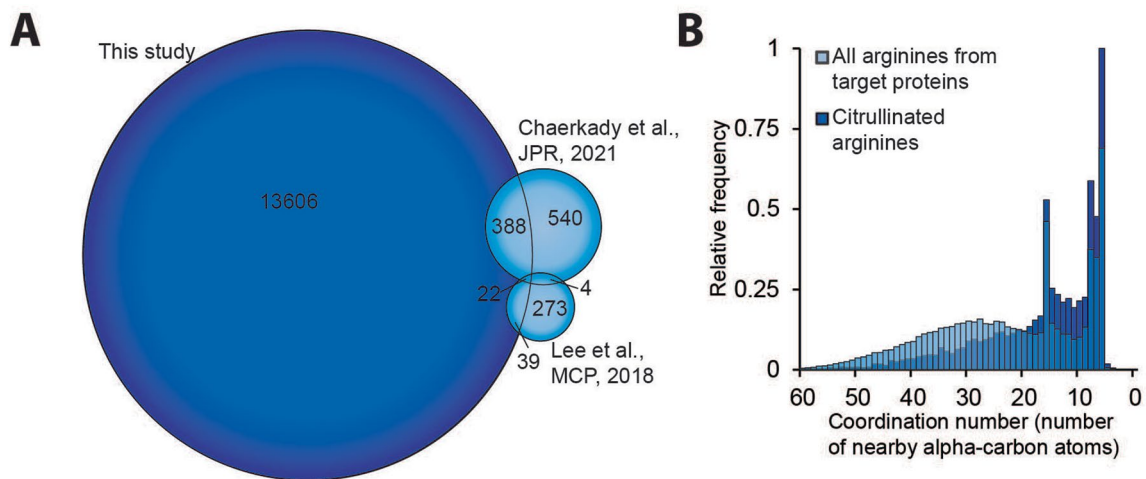
**a)** Protein copy number of proteins detected in pilot study (Exp. 1) (Figs. 1 and 2) and GSK experiment (Exp. 2) with optimized deeper sequencing. Whiskers: 95th and 5th percentile, box limits; 3rd and 1st quartiles, center bar; median.  $n = 3$  biological independent samples for Exp. 1 and  $n = 4$  biological independent samples for Exp. 2. **b)** Venn diagram displaying distribution of sites detected in pilot study (Figs. 1 and 2) performed on HF-Exactive and optimized GSK experiment performed on Exploris systems. **c)** Volcano plot of protein detected displaying little significant protein regulation in response to treatment with 20  $\mu\text{M}$  GSK484. **d)** Volcano plot of unmodified peptides from proteins displaying significant regulation of citrullination detected during treatment with 20  $\mu\text{M}$  GSK484 vs control condition. **e)** Occupancy of citrullination across the four conditions. Whiskers: 95th and 5th percentile, box limits; 3rd and 1st quartiles, center bar; median, +symbol; average. **f)** Gene Ontology (GO) term enrichment

analysis for biological processes of citrullination targeted proteins in the 20  $\mu\text{M}$  GSK484 condition, as compared control condition without inhibitor. **g)** Regulation of citrullination on enolase in response to increasing GSK484 concentrations relative to control. **h)** Regulation of citrullination on myelin basic protein in response to increasing GSK484 concentrations relative to control. **i)** Relative histone abundance in citrullinome (grey) and proteome (blue).  $n = 4$  biological independent samples, error bars represent SD. **j)** Histone N-terminal coverage by histone isoform. Citrullination site in terminal indicated (red). Missed cleavage indicated; 0 (yellow), 1 (green) and 2 (blue). **k)** Regulation of citrullination on histone 1 variant in response to increasing GSK484 concentrations relative to control. **l)** Regulation of citrullination on histone 2A in response to increasing GSK484 concentrations relative to control. **m)** Regulation of citrullination on histone 3 in response to increasing GSK484 concentrations relative to control.



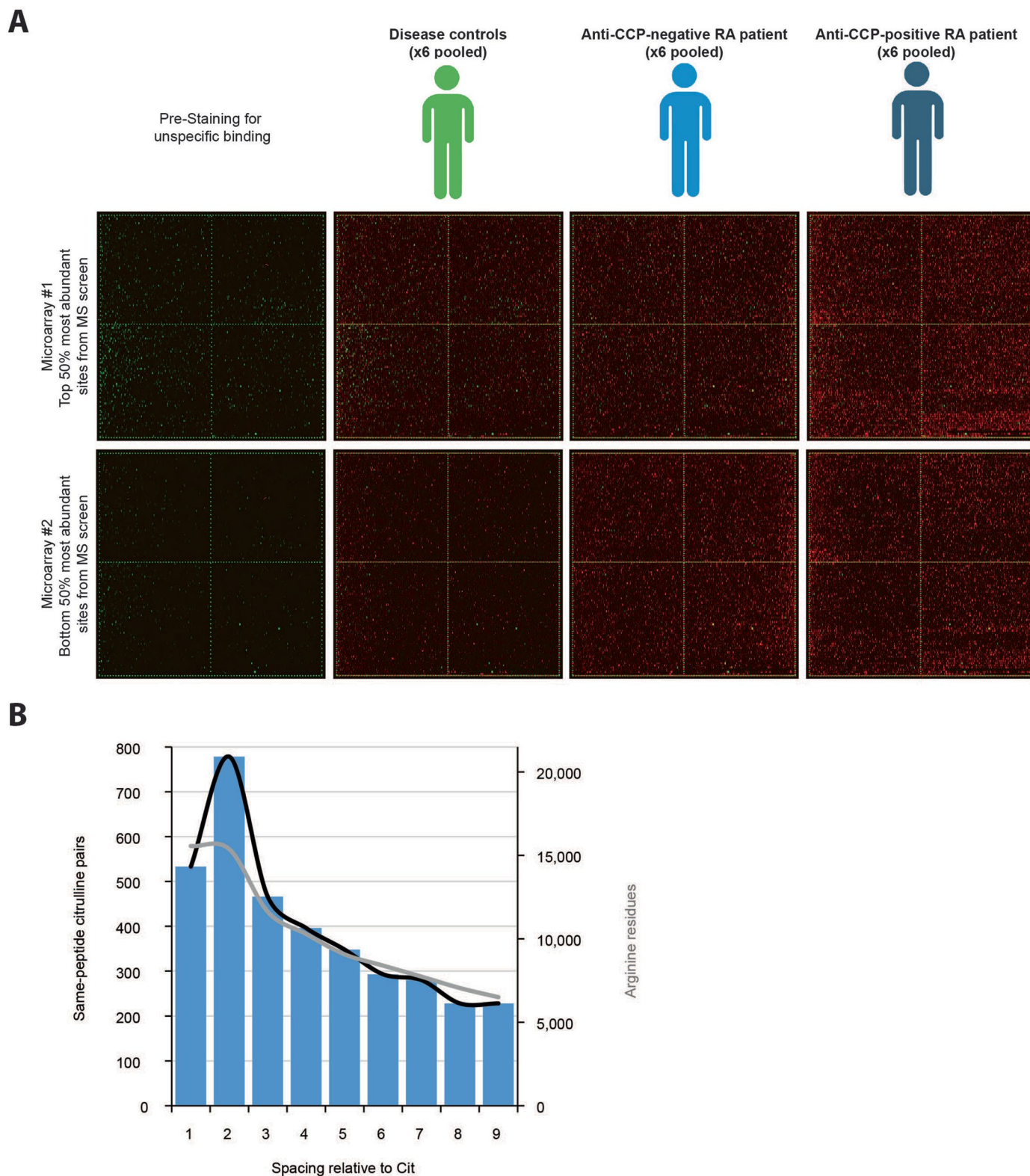
**Extended Data Fig. 4 | Further investigation of citrullination targeting.**  
**a)** Venn diagram displaying significant overlap between arginines targeted for citrullination and arginines in DNA-binding regions across all arginines in target proteins according to two-tailed Fisher Exact testing with Benjamini-Hockberg correction, enrichment factor 1.65, p-value 3.81E-07. **b)** Venn diagram displaying significant overlap between arginines targeted for citrullination and arginines in DNA-binding regions across all arginines in transcriptional regulators citrullinated by PADI4 according to two-tailed Fisher Exact testing with Benjamini-Hockberg correction, enrichment factor 1.54, p-value 0.00022.

**c)** Annotations enriched for the targets of citrullinated (top) and not citrullinated TFs (bottom) from the categories KEGG pathways, Gene Ontology Biological Process (BP), and UniProt keywords as determined by STRING gene-set enrichment analysis (see Supplementary Tables 6 and 7) for a list of all enriched annotations). Gene count refers to the number of genes annotated with the given annotation term, the false discovery rate (FDR) indicates the significance for each term, and enrichment score is the ratio between the term mean and the maximum deviation from the mean in the user input, multiplied by a factor of 10.



**Extended Data Fig. 5 | Comparison to previous studies and coordination number of arginines. a)** Scaled venn diagram displaying overall of sites detected in our study, by Chaerkady et al., 2021 and Lee et al., 2018. **b)** Histogram showing

coordination number (number of nearby alpha-carbon atoms) of citrullinated arginines and all arginines from the same target proteins. Two-tailed Spearman nonparametric correlation analysis show significant difference ( $7.53e-032$ ).



**Extended Data Fig. 6 | Raw images of peptide microarrays and peptide spacing.** a) Raw peptide microarray images of microarrays, column 1 depicts microarray #1 and #2 pre-stained with secondary antibody Goat anti-human IgG (Fc), column 2 microarrays incubated synovial fluid from disease controls,

column 3 microarrays incubated with synovial fluid from CCP- RA patients and column 4 microarrays incubated with synovial fluid from CCP+ RA patients. b) Distance of same-peptide citrulline pair spacing (black), arginine spacing indicated (grey).

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- |                                     |                                     |  |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | A description of all covariates tested   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated   |

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection No custom software was used. The MS data was collected with the software provided with the instruments.

Data analysis No custom software was used. All software used in the manuscript is publicly available and described in the methods. Software used: MaxQuant v1.5.3.30, Perseus v1.6.14.0, iceLogo v1.2, Icelogo v1.3.8, BoxPlorR web tool v1 and Sicasys v1.4 were used.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Rivero et al., 2022) partner repository with the dataset identifier PXD038702

All other data generated in this study is provided in the Supplementary Information/Source Data file. Source data are provided with this paper.

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	Sex and gender was not considered in the study design, instead patients were recruited based on needing joint-puncture and steroid treatment due to flare in the joint. A total of 18 patients were included, 12 females and 6 males. Patient ages at point of sample collection were 28, 29, 29, 31, 31, 33, 44, 49, 51, 54, 61, 62, 62, 63, 67, 72, 73, 76 years old.
Reporting on race, ethnicity, or other socially relevant groupings	There were no relevant race, ethnicity or other socially relevant categorization variable(s) used in the manuscript.
Population characteristics	The patients were divided into three groups based on diagnosis, Anti-CCP-negative rheumatoid arthritis patients, Anti-CCP-positive rheumatoid arthritis patients and patients suffering from ankylosing spondylitis. Ages ranged from 28 to 76 years old at point of sample collection.
Recruitment	Patients with RA and SpA with a clinical single joint flare where the treating physician found indication for joint puncture and aspiration - with or without glucocorticoid injection - were recruited for the project and referred for an ultrasound guided procedure by a rheumatologist related to the project thereby minimizing the self-selection bias. Patients with recent trauma to or surgery of the involved joint or who had had a glucocorticoid injection in the same joint within 3 months were excluded. None of the recruited patients declined participation in the project. We only included patients with a single joint flare and hence the inflammation level may be higher in patients with generalized flare. These patients would however never be treated with joint aspiration and injections alone. Patients were not compensated for participation.
Ethics oversight	Synovial fluid was from collected patients suffering from anti-CCP-positive RA, patients with anti-CCP-negative RA, and patients with ankylosing spondylitis, all recruited at Center for Rheumatology and Spine Diseases, Copenhagen University Hospital Glostrup after informed consent after approval by the local ethical committee (H-16042831). The local ethics committee was: De Videnskabsetiske Komiteer for Region Hovedstaden, Regionsgården Kongens Vænge 2, 3400 Hillerød, Denmark with journal nr.:H-16042831.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences  Behavioural & social sciences  Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample-size calculation was performed. MS analysis were performed in biological triplicates or quadruplicates to ensure reproducibility using HL60 cell line lysates. The size of individual samples (protein starting material) was determined during pilot experiments. The peptide microarray analysis was performed as a single replicate, n=1, due to the expansive nature of this analysis.
Data exclusions	One replicates for all conditions was excluded in study 1 of the MS analysis, due to unfortunate sample loss prior to MS analysis and this study is therefore performed in triplicates.
Replication	MS experiments were performed in triplicate or quadruplicates to ensure reproducibility as shown by principal component analysis (figure S1A) and pearson correlation (S1B). All other MS experiments similarly showed reproducibility. WB analysis was performed in at least duplicates. All attempts of replication were successful. The peptide microarray analysis was performed as a single replicate, n=1, due to this no statistical comparison between the microarrays was performed only within the same array.
Randomization	Samples were not divided into experimental groups. All replicates for all individual mass spectrometry experiments were prepared simultaneously and statistically processed while taking multiple-hypothesis testing into account.
Blinding	Sample preparation was not blinded, all samples relating to each experiment were prepared simultaneously and clearly labeled, which is important for MS analysis. During MS data acquisition, the performance of the MS instrument drifts over time or power outages and other factors outside our control may affect acquisition. Samples were therefore run in an order that introduces the least technical variance.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involvement in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

### Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

### Antibodies used

The following antibodies were used in this study: rabbit polyclonal PADI4 (cat. P4749, Sigma Aldrich), rabbit monoclonal H3 (citrulline Arg2) antibody (1:1000, cat. Ab176843, clone EPR17703, Abcam) and rabbit polyclonal GAPDH (1:1000, Ab9485, Abcam). The Anti-Citrulline (Modified) Detection Kit (cat. 17-347B, Merck) was used to measure global citrullination<sup>34</sup>, according to the manufacturer's instructions. Global citrullination was detected with the antimodified citrulline (AMC) Detection Kit (cat. 17-347B, Merck) according to the manufacturer's instructions. Goat anti-human IgG (Fc) DyLight680 was used for testing secondary binding of microarrays (0.1 µg/ml, cat. SA5-10138, Thermo Fisher Scientific).

### Validation

Anti-PADI4 (P4749) is according to Sigma Aldrich validated for WB. H3Cit (Ab176843) is validated by Abcam for WB, the same applied for anti-GAPDH (Ab9485). Thermo Fisher Scientific states that the Goat anti-human IgG (SA5-10138) is validated for immunoassays.

## Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

### Cell line source(s)

HL60 cell line derived from peripheral blood lymphocytes of a women suffering from acute promyelocytic leukaemia. The HL60 cell line was gifted to us by Maria Christophorou.

### Authentication

The cell line was not authenticated

### Mycoplasma contamination

The cell lines were not tested for mycoplasma contamination

### Commonly misidentified lines (See [ICLAC](#) register)

No commonly misidentified cell lines were used in this study.