**Identification of PKC-regulated phosphosites on LRRK1 by mass spectrometry analysis**

**Authors: Asad U Malik, Raja S. Nirujogi, Toan K. Phung, Dario R. Alessi**

Medical Research Council Protein Phosphorylation and Ubiquitylation Unit, School of Life Sciences, University of Dundee, Dow Street, Dundee DD1 5EH, UK

**Summary**

We describe a non-radioactive, mass spectrometry-based assay that we deploy for identifying novel PKC-regulated sites on LRRK1 that are responsible for activation of its kinase activity.

**1) MATERIALS**

**1.1) Reagents:**

1.1.1) Recombinant PKC protein (available from MRC Reagents and Services: <https://mrcppureagents.dundee.ac.uk/>)

1.1.2) Recombinant LRRK1 wild type [and kinase inactive? D1409A, 27-2015] protein

*Note:* Recombinant LRRK1 protein is expressed and purified by following the protocol described in: **XXXXX**

1.1.3) Kinase assay buffer: 25 mM HEPES pH 7.5; 0.1% (v/v) 2-mercaptoethanol; 50 mM KCl; 1 mM CaCl2; 10 mM MgCl2; 1 mM ATP

1.1.4) L-α-Phosphatidylserine (Avanti Polar Lipids, resuspended in methanol and chloroform at a 1:1 ratio for long-term storage)

1.1.5) L-α-Diacylglyerol (Avanti Polar Lipids, resuspended in methanol and chloroform at a 1:1 ratio for long-term storage)

1.1.6) 4X Loading buffer: Invitrogen™ NuPAGE™ LDS Sample Buffer, cat no NP0007, or 4X SDS loading buffer: 250mM Tris-HCl, pH6.8, 8% (w/v) SDS, 40% (v/v) glycerol, 0.02% (w/v) bromophenol blue.

1.1.7) SDS-PAGE buffer: For NuPAGETM Bis-Tris gels: NuPAGE MOPS SDS running buffer (ThermoFisherScientific, Cat#NP000102); for self-cast Bis-Tris gels: 50 mM MOPS, 50 mM Tris, 0.1% (w/v) SDS, 1 mM EDTA.

1.1.8) InstantBlue® Coomassie Protein Stain (abcam ab119211, or equivalent)

1.1.9) Dithiothreitol (Sigma #43815)

1.1.10) Ammonium bicarbonate #A6141

1.1.11) LC-MS grade Acetonitrile (VWR #1.00030.2500)

1.1.12) Iodoacetamide (Sigma #I1149)

1.1.13) Trifluroacetic acid (Sigma #302031)

*Note: Prepare a 20% (by vol) aqueous* trifluroacetic acid (*TFA) stock and store at 4 oC.*

1.1.14) Sequencing grade trypsin (5 X 20 ug pack. Promega #V5111).

1.1.15) Chymotrypsin (Promega # V1061 1X 25 µg)

1.1.16) Asp-N (Promega # V1621 1X 2 µg)

*Note: Store protease stocks at -20oC and thaw on ice? just before the digestion step.*

**1.2) Equipment:**

1.2.1) Refrigerated bench-top centrifuge (Eppendorf microcentrifuge 5417R, or equivalent).

1.2.2) Savant SpeedVac system (Thermo #SPD140DDA, or equivalent)

1.2.3) Thermo mixer (Eppendorf ThermoMixer, or equivalent)

1.2.4) Disposable Glass Culture Tubes (Fisherbrand Round Bottom Disposable Borosilicate Glass Tubes, or equivalent)

1.2.5) XCell4 SureLock Midi-Cell Electrophoresis System (if using Invitrogen NuPAGE precast midi gels), or equivalent gel electrophoresis apparatus.

1.2.6) See-saw rocker (VWR SSL4, or equivalent).

1.2.7) 1.5 ml protein low bind eppendorf tubes (Eppendorf™ #022431081)

1.2.8) 16-gauge needle (# Z261378. Sigma Aldrich)

1.2.9) Spray duster (Qconnect #KFO4499)

1.2.10) PTFE-O rings

*Note: Place the PTFE-O-ring on top of the Eppendorf tube to serve as an adaptor such a way that 3/4th of the Stage-tip could be placed into the tube during the centrifugation step. PTFE-O-rings can be purchased from NEST group desalting columns and re use them* [*https://www.nestgrp.com/*](https://www.nestgrp.com/)

1.2.11) X72 40 mL Amber class EPA vial W Cap and seal (Cole Parmer # 10572553)

1.2.12) Empore C18 disks, 47 mm (CDS analytical #2215)

*Note: Prepare a single layer with 16-gauge needle and pass it with spray duster into the 250 µl tip for 0.1 to 5 ug of peptide amount. For more than 5 ug, punch 2 or 3 layers with 16-guage needle.*

1.2.13) Exploris 240 Mass spectrometer

1.2.14) EvoSep Liquid chromatrography system

Note: Any nano-LC such as Easy nLC or Ultimate 3000 Dionex can be used instead.

1.2.15) Proteome Discoverer 2.4 software suite with SEQUEST or Mascot search algorithm

**2) METHOD:**

**2.1) Preparation of lipid vesicles for PKC activation**

2.1.1) Clean a disposable glass culture tube by washing three times with 100% methanol. Allow to air-dry.

2.1.2) Pipette 0.5 μl of Diacylglycerol (stock concentration is 10 mg/ml) and 5 μl of Phosphatidylserine (stock concentration is 10 mg/ml) into the cleaned and dried glass tube.

*Note: These quantities will provide sufficient lipid vesicles for 25 reactions at a volume of 20 μl per reaction.*

2.1.3) Vacuum dry lipids using a SpeedVac system for 10 minutes. This should leave a visible, translucent lipid pellet.

*Note: Ensure that lipids are completely dried as any residual chloroform or methanol will inhibit the kinase reaction.*

2.1.4) Resuspend lipids from step 2.1.3 in 50 μl of 25 mM HEPES pH 7.4, 50 mM KCl. Vortex gently until pellet is no longer visible.

**2.2) Kinase Reaction: Phosphorylation of LRRK1 by PKC**

2.2.1) Prepare a primary “2X master mix” containing 50 mM HEPES pH 7.5, 100 mM KCl, 0.2% (v/v) 2‐Mercaptoethanol, 20 mM MgCl2, 2 mM ATP, 2 mM CaCl2, 200 μg/ml Phosphatidylserine and 20 μg/ml Diacylglycerol.

2.2.2) For each reaction, add 15 μl of the primary “2X master mix” to a clean Eppendorf tube.

2.2.3) Add 7.5 μl of 200 nM LRRK1 wild type protein (final concentration is 50 nM) to each reaction and allow equilibration on ice for 5 minutes.

2.2.4) Start the kinase reaction by adding 7.5 μl of 400 nM PKC Alpha protein (final concentration is 100 nM).

*Note: The final reaction volume should be* 3*0 μl.*

*Note: Reactions not including PKC Alpha are also included as a negative control to identify phosphorylation sites that are only present when recombinant LRRK1 protein is incubated with PKC Alpha. In these reactions, add 7.5 μl of 25 mM HEPES pH 7.4, 50 mM KCl instead of PKC Alpha protein.*

2.2.5) Transfer the Eppendorf tubes to the thermo mixer set at 30oC, 1,000 rpm. Incubate for 45 minutes.

2.2.6) Stop the kinase reaction by adding 10 μl of 4X LDS loading buffer to the reaction mix to a final concentration of 1X.

2.2.7) Incubate the samples for 5 min at 70°C on a heat block before proceeding to step 2.3.

**2.3) SDS-polyacrylamide gel electrophoresis (SDS-PAGE):**

2.3.1) Load samples onto a NuPAGE 4–12% Bis–Tris Midi Gel (ThermoFisherScientific, Cat#WG1402BOX or Cat#WG1403BOX), alongside pre-stained molecular weight markers (ranging from 10 kDa to 250 kDa). Rinse wells carefully with running buffer before loading samples.

Note: Load the complete reaction onto gels to ensure detection of proteins by Instant Blue stain

2.3.2) Electrophorese samples at 130V with MOPS SDS running buffer for 2 hours or until the blue dye runs off the gel.

2.3.3) Place gel in a clean glass 15 cm dish and cover with 15-20 ml of InstantBlue® Coomassie Protein stain. Incubate on see-saw rocker for 1 hour at room temperature.

2.3.4) Replace the InstantBlue® Protein stain with double distilled water and allow to de-stain at room temperature overnight before proceeding with peptide digestion as described in 2.4.

**2.4) Total Protein Digestion**

2.4.1) Using a clean scalpel, excise stained-bands corresponding to LRRK1 from gel and cut into approximately 1mm2 gel pieces.

2.4.2) Transfer the gel pieces into a low-bind tube.

2.4.3) De-stain gel pieces by repeated 10 min washes in 40% (v/v) ACN in 40mM NH4HCO3.

*Note: Wash by incubation on thermomixer set to 1,200 rpm at room temperature. Repeat step 2.4.3 until gel pieces are completely colourless.*

2.4.4) Reduce peptides by addition of 100 µl of 5 mM DTT in 40 mM NH4HCO3. Incubate on thermomixer at 56 oC for 30 mins, 1200 rpm.

2.4.5) Remove the DTT solution and incubate gel pieces in 40% (v/v) ACN in 40mM NH4HCO3 for 10 minutes at room temperature?

*Note: This step allows the gel pieces to subsequently imbibe iodoacetamide (Step 2.4.6).*

2.4.6) Alkylate peptides by addition of 20mM iodoacetamide in 40mM NH4HCO3 and incubate at room temperature for 30 minutes, 1,200 rpm.

*Note: Samples should be kept in the dark during this step as iodoacetamide is light-sensitive.*

2.4.7) Dehydrate gel pieces by washing in 100% (v/v) ACN for 10 min

*Note: Perform this step on thermomixer set to 1,200 rpm at room temperature. Repeat step 2.4.7 twice until the gel pieces appear completely dry and white.*

2.4.8) Remove supernatant using a pipette and vacuum dry gel pieces to remove any residual CAN.

2.4.9) Add 100ng of protease in 100ul of appropriate buffer (See Table 1) to the gel pieces from step 2.4.8 and incubate overnight on thermomixer at 37 oC, 1,200 rpm.

Note: Table 1 describes the different protease combinations used for total protein digestion and the appropriate buffers for each protease.

|  |  |
| --- | --- |
| **Protease** | **Buffer** |
| Trypsin + LysC | 50 mM TEABC |
| Asp-N | 50 mM Tris-HCl |
| Chymotrypsin | 100 mM Tris-HCl + 10 mM CaCl2 |

Table 1: Protease combinations used for total protein digestion and appropriate buffers for each protease.

**2.5) Peptide extraction**

2.5.1) Supplement samples from step 2.4.9 with 50ul of extraction buffer (80% ACN in 0.2% Formic Acid) and incubate on thermomixer at room temperature for 10 min at 1,200 rpm.

2.5.2) Centrifuge samples for 1 min at 2000g to pellet the gel pieces and using a pipette carefully transfer the supernatant to a new low-binding? tube.

*Note: Ensure that the gel pieces are not transferred to the new tube when pipetting the supernatant.*

2.5.3) Repeat step 2.5.1 until the gel pieces appear completely dried. Each time, transfer the supernatant into the same tube (from step 2.5.2).

2.5.4) Vacuum dry the combined supernatants (containing the digested peptides) and proceed with C18 clean-up protocol (as described in 2.6).

**2.6) C18 stage-tip protocol:**

*Note:This protocol has been adapted from* [*dx.doi.org/10.17504/protocols.io.bs3tngnn*](https://dx.doi.org/10.17504/protocols.io.bs3tngnn)

2.6.1) Prepare single layer of C18 stage-tip using 16-gauge syringe needle.

*Note: Prepare a single layer with 16-gauge needle and pass it with spray duster into the 250 µl tip for 0.1 to 5 ug of peptide amount.*

2.6.2) Resuspend the vacuum dried peptides from step 2.5.4 in 80 µl of Solvent A1 (0.1% (by vol) TFA in MQ-H2O).

2.6.3) Add 80 µl of 100% (by vol) ACN to the C18 stage-tip from Step 2.6.1 and centrifuge at 2000g for 2 min at room temperature. Discard flow through.

*Note: This step is required to activate the C18 resin.*

2.6.4) Add 80 µl Solvent A1 (0.1% (by vol) TFA (by vol) in MQ-H2O)) and centrifuge at 2000g for 2 min at room temperature. Discard flow through. Repeat this step.

*Note: This step is required to equilibrate the C18 resin.*

2.6.5) Load the acidified peptide digest from Step 2.6.2 to the C18 stage-tip from step 2.6.4 and centrifuge at 1500g for 5 min at room temperature.

*Note: During this step the peptides will absorb to the C18 resin.*

2.6.6) Reapply the flow through to the C18 stage-tip column and centrifuge at 1500g for 5 min at room temperature.

2.6.7) Add 80 µl of Solvent A1 (0.1% (by vol) TFA v/v) in MQ-H2O)?) to the C18 stage-tip column and centrifuge at 2,000g for 2 min at room temperature. Discard flow through. Repeat again.

2.6.8) Place the C18 stage-tip from step 2.6.7 into a new 1.5 ml low binding tube.

*Note: Using new tubes is important to avoid contamination.*

2.6.9) Elute peptides from the C18 stage-tip by adding 40 µl of Elution buffer (Solvent B1: 40% (by vol) acetonitrile in 0.1% (by vol) TFA) in MQ-H2O and centrifuge at 1,500g for 2 min.

2.6.10) Repeat step 2.6.9.

2.6.11) Immediately snap freeze the eluted peptides from step 2.6.10 on dry ice and vacuum dry.

2.6.12) Perform mass spectrometry analysis of the peptides as described in 2.7.

2.7) LC-MS/MS analysis

2.7.1) Dissolve the peptides in LC-Buffer (3% ACN (v/v) in 0.1% Formic acid (v/v)). (Note: Just 200ng of peptide digest per sample is good enough to achieve the coverage on Exploris 240 mass spectrometer. If the starting material of LRRK1 that was used for the Kinase assay is 1µg then split the sample into five aliquots of 200ng each to inject on MS. The reminder of the sample can be injected on a different mass spectrometer to get an alternative fragmentation to HCD such as EThCD on Lumos or EAD on Sciex Zeno-TOF 7600 MS platforms).

2.7.2) Take 200ng of the peptide digest of LRRK2 in 5 or 10µl in LC-buffer and prepare it for the Evotips loading. The Evo tips are a versatile disposable trap columns that enables <0.1% carry-over between samples.

2.7.3) Prepare the Evotips as described in the Protocol in PMID: 33367571.

2.7.4) Place the Evotips on EvoSep autosampler and used the 30 sample per day (30SPD) method to execute the LC method through Xcalibur interface that is inline with Orbitrap Exploris 240 mass spectrometer.

2.7.5) EvoSep LC system injects and executes a partial elution of the sample from Evotip and loads onto the long storage loop in which the pre-formed gradient generated at the initial step. Following the loading the High-pressure pump pushes the sample into the analytical column (ReproSil-Pur C18, 1.9 µm beads by Dr Maisch. #EV1113).

2.7.6) The following MS instrument method can be constructed for the High-resolution HCD fragmentation analysis:

|  |  |  |
| --- | --- | --- |
| Instrument | Thermo Scientific Orbitrap Exploris 240 |  |
| LC system | EvoSep Liquid Chromatography system | 30 SPD method |
| Method duration | 45 min |  |
| MS Global settings: |  |  |
|  | Infusion mode: | Liquid Chromatography |
|  | Expected LC peak width (s): | 15 |
|  | Advanced Peak determination: | TRUE |
|  | Default charge state: | 2 |
|  | Internal mass calibration: | off |
|  |  |  |
| Full scan settings: |  |  |
|  | Orbitrap resolution: | 120000 |
|  | Scan range (m/z): | 375-1500 |
|  | RF lens(%): | 70 |
|  | AGC target: | Custom |
|  | Normalized AGC target (%): | 300 |
|  | Maximum injection Time mode: | Custom |
|  | Maximum injection Time (ms): | 25 |
|  | Micorscans: | 1 |
|  | Data type: | Profile |
|  | Polarity: | Positive |
| Filters: |  |  |
| MIPS | Monoisotopic peak determination: | Peptide |
|  | Relax restrictions when too few precursors are found: | TRUE |
| Intensity | Filter Type: | Intensity Threshold |
|  | Intensity Threshold: | 5.00E+03 |
| Charge State | Include charge state(s): | 2 to 6 |
|  | Include undetermined charge states: | False |
| Dynamic Exclusion | Dynamic Exclusion Mode: | Custom |
|  | Exclude after n times: | 1 |
|  | Exclusion duration (s): | 5 |
|  | Mass Tolerance: | ppm |
|  | Low: | 10 |
|  | High | 10 |
|  | Exclude isotopes: | TRUE |
|  | Perform dependent scan on single charge state per precursor only: | FALSE |
| Data Dependent | Data Dependent Mode: | Number of Scans |
|  | Number of Dependent Scans | 10 |
| ddMS2 settings | Isolation Window (m/z): | 1.2 |
|  | Isolation Offset: | Off |
|  | Collision Energy Mode: | Fixed |
|  | Collision Energy Type: | Normalized |
|  | HCD Collision Energy (%): | 28 |
|  | Orbitrap resolution: | 15000 |
|  | First Mass (m/z): | 110 |
|  | Scan range mode: | Auto |
|  | AGC target: | Standard |
|  | Maximum injection Time mode: | Custom |
|  | Maximum injection Time (ms): | 100 |
|  | Micorscans: | 1 |
|  | Data type: | Profile |
|  | Polarity: | Positive |

2.8) Data analysis:

2.8.1) Transfer the raw data to search with Thermo Scientific Proteome Discoverer 2.4 Software suite that is integrated with Sequest-HT search algorithm (Optional: As the PD 2.4 software is commercial software suite, if you don’t have access to it consider in using Open-source package like MaxQuant or FragPipe.)

2.8.2) We recommend creating a custom protein sequence FASTA file rather than using the entire Uniprot Human or Mouse proteome FASTA file. For example: Copy the Human LRRK1 FASTA sequence and past it into a Notepad++ and save with LRRK1.FASTA (Note: Ensure if you have any N-ter or C-ter GFP or HA tag of a recombinant LRRK1 and append the sequence accordingly).

2.8.3) Import the LRRK1.FASTA sequence into the PD 2.4 software

2.8.4) Construct the Processing and Consensus workflows

|  |  |  |
| --- | --- | --- |
| ------------------------------------------------------------------ |  |  |
| The Processing workflow tree |  |  |
| ------------------------------------------------------------------ |  |  |
|  |  |  |
| (0) Spectrum Files |  |  |
| (1) Spectrum Selector |  |  |
| (2) Sequest HT |  |  |
| (3) Fixed Value PSM Validator |  |  |
| (4) IMP-ptmRS |  |  |
| (5) Minora Feature Detector |  |  |
|  |  |  |
| ------------------------------------------------------------------ |  |  |
| Processing node 0 | Spectrum Files |  |
| ------------------------------------------------------------------ |  |  |
| Input Data |  | Note |
| File Name(s) |  | Specify the sample condtion and the Enyzme associated with the digestion |
|  | RN-AM\_211216\_LRRK1\_+PKC\_Tryp-LysC\_01.raw |  |
|  | RN-AM\_211216\_LRRK1\_+PKC\_Tryp-LysC\_01.raw |  |
|  | RN-AM\_211216\_LRRK1\_-PKC\_Tryp-LysC\_01.raw |  |
|  | RN-AM\_211216\_LRRK1\_-PKC\_Tryp-LysC\_01.raw |  |
|  |  |  |
| ------------------------------------------------------------------ |  |  |
| Processing node 1 | Spectrum Selector |  |
| ------------------------------------------------------------------ |  |  |
| 1. General Settings |  |  |
| Precursor Selection | Use MS1 Precursor |  |
| Use Isotope Pattern in Precursor Reevaluation | True |  |
| Provide Profile Spectra | Automatic |  |
|  |  |  |
| 2. Spectrum Properties Filter |  |  |
| Lower RT Limit | 0 |  |
| Upper RT Limit | 0 |  |
| First Scan | 0 |  |
| Last Scan | 0 |  |
| Lowest Charge State | 0 |  |
| Highest Charge State | 0 |  |
| Min. Precursor Mass | 350 Da |  |
| Max. Precursor Mass | 5000 Da |  |
| Total Intensity Threshold | 0 |  |
| Minimum Peak Count | 1 |  |
|  |  |  |
| 3. Scan Event Filters |  |  |
| Mass Analyzer | Is FTMS |  |
| MS Order | Is MS2; MS1 |  |
| Activation Type | Is HCD |  |
| Min. Collision Energy | 0 |  |
| Max. Collision Energy | 1000 |  |
| Scan Type | Is Full |  |
| Polarity Mode | Is + |  |
|  |  |  |
| 4. Peak Filters |  |  |
| - S/N Threshold (FT-only) | 1.5 |  |
|  |  |  |
| 5. Replacements for Unrecognized Properties |  |  |
| Unrecognized Charge Replacements | Automatic |  |
| Unrecognized Mass Analyzer Replacements | FTMS |  |
| Unrecognized MS Order Replacements | MS2 |  |
| Unrecognized Activation Type Replacements | HCD |  |
| Unrecognized Polarity Replacements | + |  |
| Unrecognized MS Resolution@200 Replacements | 120000 |  |
| Unrecognized MSn Resolution@200 Replacements | 30000 |  |
|  |  |  |
| 6. Precursor Pattern Extraction |  |  |
| Precursor Clipping Range Before | 2.5 Da |  |
|  | 5.5 Da |  |
|  |  |  |
| ------------------------------------------------------------------ |  |  |
| Processing node 2 | Sequest HT |  |
| ------------------------------------------------------------------ |  |  |
| 1. Input Data |  |  |
| Protein Database | LRRK1.FASTA |  |
| Enzyme Name | Trypsin (Full) | Here, specify AspN and Chymotrypsin separately fof the searches associated with those conditions |
| Max. Missed Cleavage Sites | 2 |  |
| Min. Peptide Length | 7 |  |
| Max. Peptide Length | 144 |  |
| Max. Number of Peptides Reported | 10 |  |
|  |  |  |
| 2. Tolerances |  |  |
| Precursor Mass Tolerance | 10 ppm |  |
| Fragment Mass Tolerance | 0.05 Da |  |
| Use Average Precursor Mass | False |  |
| Use Average Fragment Mass | False |  |
|  |  |  |
| 3. Spectrum Matching |  |  |
| Use Neutral Loss a Ions | True |  |
| Use Neutral Loss b Ions | True |  |
| Use Neutral Loss y Ions | True |  |
| Use Flanking Ions | True |  |
| Weight of a Ions | 0 |  |
| Weight of b Ions | 1 |  |
| - Weight of c Ions | 0 |  |
| Weight of x Ions | 0 |  |
| Weight of y Ions | 1 |  |
| Weight of z Ions | 0 |  |
|  |  |  |
| 4. Dynamic Modifications |  |  |
| Max. Equal Modifications Per Peptide | 3 |  |
| Max. Dynamic Modifications Per Peptide | 4 |  |
| - 1. Dynamic Modification | Oxidation / +15.995 Da (M) |  |
| - 2. Dynamic Modification | Phospho / +79.966 Da (S, T, Y) |  |
|  |  |  |
| 7. Static Modifications |  |  |
| - 1. Static Modification | Carbamidomethyl / +57.021 Da (C) |  |
|  |  |  |
| ------------------------------------------------------------------ |  |  |
| Processing node 3 | Fixed Value PSM Validator |  |
| ------------------------------------------------------------------ |  |  |
| 1. Input Data |  |  |
| Maximum Delta Cn | 0.05 |  |
| Maximum Rank | 0 |  |
|  |  |  |
| ------------------------------------------------------------------ |  |  |
| Processing node 4 | IMP-ptmRS |  |
| ------------------------------------------------------------------ |  |  |
| 1. Scoring |  |  |
| PhosphoRS Mode | True |  |
| Report only PTMs | True |  |
| Use Diagnostic Ions | True |  |
| Use Fragment Mass Tolerance of Search Node | True |  |
| Fragment Mass Tolerance | 0.5 Da |  |
| Consider Neutral Loss peaks for CID, HCD and EThcD | Automatic |  |
| Maximum Peak Depth | 8 |  |
| Use a Mass accuracy correction | False |  |
|  |  |  |
| 2. Performance |  |  |
| Maximum Number of Position Isoforms | 500 |  |
| Maximum PTMs Per Peptide | 10 |  |
|  |  |  |
| ------------------------------------------------------------------ |  |  |
| Processing node 5 | Minora Feature Detector |  |
| ------------------------------------------------------------------ |  |  |
| 1. Peak & Feature Detection |  |  |
| Min. Trace Length | 5 |  |
| - Max. ΔRT of Isotope Pattern Multiplets [min] | 0.2 |  |
|  |  |  |
| 2. Feature to ID Linking |  |  |
| PSM Confidence At Least | High |  |

|  |  |
| --- | --- |
| The Consensus workflow tree |  |
| ------------------------------------------------------------------ |  |
|  |  |
| (0) MSF Files |  |
| (1) PSM Grouper |  |
| (2) Peptide Validator |  |
| (3) Peptide and Protein Filter |  |
| (4) Protein Scorer |  |
| (5) Protein Grouping |  |
| (6) Peptide in Protein Annotation |  |
| (15) Modification Sites |  |
| (7) Protein FDR Validator |  |
| (16) Peptide Isoform Grouper |  |
| (10) Feature Mapper |  |
| (11) Precursor Ions Quantifier |  |
|  |  |
| Post-processing nodes |  |
| -------------------------------- |  |
|  |  |
| (12) Result Statistics |  |
| (13) Display Settings |  |
| (14) Data Distributions |  |
|  |  |
| ------------------------------------------------------------------ |  |
| Processing node 0 | MSF Files |
| ------------------------------------------------------------------ |  |
| 1. Storage Settings |  |
| Spectra to Store | Identified or Quantified |
| Feature Traces to Store | All |
|  |  |
| 2. Merging of Identified Peptide and Proteins |  |
| Merge Mode | Globally by Search Engine Type |
|  |  |
| 3. FASTA Title Line Display |  |
| Reported FASTA Title Lines | Best match |
| Title Line Rule | standard |
|  |  |
| 4. PSM Filters |  |
| Maximum Delta Cn | 0.05 |
| Maximum Rank | 0 |
| Maximum Delta Mass | 0 ppm |
|  |  |
| Hidden Parameters |  |
| MSF File(s) | RN-AM\_211216\_LRRK1\_Sequest-Trypsin-(1).msf |
|  |  |
| ------------------------------------------------------------------ |  |
| Processing node 1 | PSM Grouper |
| ------------------------------------------------------------------ |  |
| 1. Peptide Group Modifications |  |
| Site Probability Threshold | 75 |
|  |  |
| ------------------------------------------------------------------ |  |
| Processing node 2 | Peptide Validator |
| ------------------------------------------------------------------ |  |
| 1. General Validation Settings |  |
| Validation Mode | Automatic (Control peptide level error rate if possible) |
| Target FDR (Strict) for PSMs | 0.01 |
| Target FDR (Relaxed) for PSMs | 0.05 |
| Target FDR (Strict) for Peptides | 0.01 |
| Target FDR (Relaxed) for Peptides | 0.05 |
|  |  |
| 2. Specific Validation Settings |  |
| Validation Based on | q-Value |
| Target/Decoy Selection for PSM Level FDR Calculation Based on Score | Automatic |
| Reset Confidences for Nodes without Decoy Search (Fixed Score thresholds) | False |
|  |  |
| ------------------------------------------------------------------ |  |
| Processing node 3 | Peptide and Protein Filter |
| ------------------------------------------------------------------ |  |
| 1. Peptide Filters |  |
| Peptide Confidence At Least | High |
| Keep Lower Confident PSMs | False |
| Minimum Peptide Length | 7 |
| Remove Peptides without Protein Reference | False |
|  |  |
| 2. Protein Filters |  |
| Minimum Number of Peptide Sequences | 1 |
| Count Only Rank 1 Peptides | False |
| Count Peptides only for Top Scored Protein | False |
|  |  |
| ------------------------------------------------------------------ |  |
| Processing node 4 | Protein Scorer |
| ------------------------------------------------------------------ |  |
| No parameters |  |
|  |  |
| ------------------------------------------------------------------ |  |
| Processing node 5 | Protein Grouping |
| ------------------------------------------------------------------ |  |
| 1. Protein Grouping |  |
| Apply Strict parsimony principle | True |
|  |  |
| ------------------------------------------------------------------ |  |
| Processing node 6 | Peptide in Protein Annotation |
| ------------------------------------------------------------------ |  |
| 1. Flanking Residues |  |
| Annotate Flanking Residues of the Peptide | True |
| Number Flanking Residues in Connection Tables | 1 |
|  |  |
| 2. Modifications in Peptide |  |
| Protein Modifications Reported | Only for Master Proteins |
|  |  |
| 3. Modifications in Protein |  |
| Modification Sites Reported | All And Specific |
| Minimum PSM Confidence | High |
| Report only PTMs | True |
|  |  |
| 4. Positions in Protein |  |
| Protein Positions for Peptides | Only for Master Proteins |
|  |  |
| ------------------------------------------------------------------ |  |
| Processing node 15 | Modification Sites |
| ------------------------------------------------------------------ |  |
| 1. General |  |
| Report only PTMs | True |
| only Master Proteins | True |
| Motif Radius | 10 |
|  |  |
| ------------------------------------------------------------------ |  |
| Processing node 7 | Protein FDR Validator |
| ------------------------------------------------------------------ |  |
| 1. Confidence Thresholds |  |
| Target FDR (Strict) | 0.01 |
| Target FDR (Relaxed) | 0.05 |
|  |  |
| ------------------------------------------------------------------ |  |
| Processing node 16 | Peptide Isoform Grouper |
| ------------------------------------------------------------------ |  |
| No parameters |  |
|  |  |
| ------------------------------------------------------------------ |  |
| Processing node 10 | Feature Mapper |
| ------------------------------------------------------------------ |  |
| 1. Chromatographic Alignment |  |
| Perform RT Alignment | True |
| - Maximum RT Shift [min] | 10 |
| Mass Tolerance | 10 ppm |
| Parameter Tuning | Coarse |
|  |  |
| 2. Feature Linking and Mapping |  |
| RT Tolerance [min] | 0 |
| Mass Tolerance | 0 ppm |
| Min. s/N Threshold | 5 |
|  |  |
| ------------------------------------------------------------------ |  |
| Processing node 11 | Precursor Ions Quantifier |
| ------------------------------------------------------------------ |  |
| 1. General Quantification Settings |  |
| Peptides to Use | Unique + Razor |
| Consider Protein Groups for Peptide Uniqueness | True |
| Use Shared Quan Results | True |
| Reject Quan Results with Missing Channels | False |
|  |  |
| 2. Precursor Quantification |  |
| Precursor Abundance Based on | Intensity |
| Min. # Replicate Features [%] | 0 |
|  |  |
| 3. Normalization and Scaling |  |
| Normalization Mode | Total Peptide Amount |
| Scaling Mode | On All Average |
|  |  |
| 4. Exclude Peptides from Protein Quantification |  |
| for Normalization | Use All Peptides |
| for Protein Roll-Up | Use All Peptides |
| for Pairwise Ratios | Exclude Modified |
|  |  |
| 5. Quan Rollup and Hypothesis Testing |  |
| Protein Abundance Calculation | Summed Abundances |
| N for Top N | 3 |
| Protein Ratio Calculation | Pairwise Ratio Based |
| Maximum Allowed Fold Change | 100 |
| Imputation Mode | None |
| Hypothesis Test | t-test (Background Based) |
|  |  |
| 6. Quan Ratio Distributions |  |
| - 1st Fold Change Threshold | 2 |
| - 2nd Fold Change Threshold | 4 |
| - 3rd Fold Change Threshold | 6 |
| - 4th Fold Change Threshold | 8 |
| - 5th Fold Change Threshold | 10 |

2.8.5) If the database search is to be done using MaxQuant then refer below settings

|  |  |
| --- | --- |
| Parameter | Value |
| Version | 2.0.3.0 |
| User name | RNirujogi |
| Machine name | MRC-MS-R640-4 |
| Date of writing | 05/23/2022 15:15:41 |
| Include contaminants | TRUE |
| PSM FDR | 0.01 |
| SM FDR Crosslink | 0.01 |
| Protein FDR | 0.01 |
| Site FDR | 0.01 |
| Use Normalized Ratios For Occupancy | TRUE |
| Min. peptide Length | 7 |
| Min. score for unmodified peptides | 0 |
| Min. score for modified peptides | 40 |
| Min. delta score for unmodified peptides | 0 |
| Min. delta score for modified peptides | 6 |
| Min. unique peptides | 0 |
| Min. razor peptides | 1 |
| Min. peptides | 1 |
| Use only unmodified peptides and | TRUE |
| Modifications included in protein quantification | Oxidation (M);Acetyl (Protein N-term);Deamidation (NQ) |
| Peptides used for protein quantification | Razor |
| Discard unmodified counterpart peptides | TRUE |
| Label min. ratio count | 2 |
| Use delta score | FALSE |
| iBAQ | FALSE |
| iBAQ log fit | FALSE |
| Match between runs | FALSE |
| Find dependent peptides | FALSE |
| Fasta file | C:\Raja\Database\LRRK1.FASTA |
| Decoy mode | revert |
| Include contaminants | TRUE |
| Advanced ratios | TRUE |
| Fixed andromeda index folder |  |
| Combined folder location |  |
| Second peptides | TRUE |
| Stabilize large LFQ ratios | TRUE |
| Separate LFQ in parameter groups | FALSE |
| Require MS/MS for LFQ comparisons | TRUE |
| Calculate peak properties | FALSE |
| Main search max. combinations | 200 |
| Advanced site intensities | TRUE |
| Write msScans table | FALSE |
| Write msmsScans table | TRUE |
| Write ms3Scans table | TRUE |
| Write allPeptides table | TRUE |
| Write mzRange table | TRUE |
| Write DIA fragments table | FALSE |
| Write DIA fragments quant table | FALSE |
| Write pasefMsmsScans table | TRUE |
| Write accumulatedMsmsScans table | TRUE |
| Max. peptide mass [Da] | 4600 |
| Min. peptide length for unspecific search | 8 |
| Max. peptide length for unspecific search | 25 |
| Razor protein FDR | TRUE |
| Disable MD5 | FALSE |
| Max mods in site table | 3 |
| Match unidentified features | FALSE |
| Epsilon score for mutations |  |
| Evaluate variant peptides separately | TRUE |
| Variation mode | None |
| MS/MS tol. (FTMS) | 20 ppm |
| Top MS/MS peaks per Da interval. (FTMS) | 12 |
| Da interval. (FTMS) | 100 |
| MS/MS deisotoping (FTMS) | TRUE |
| MS/MS deisotoping tolerance (FTMS) | 7 |
| MS/MS deisotoping tolerance unit (FTMS) | ppm |
| MS/MS higher charges (FTMS) | TRUE |
| MS/MS water loss (FTMS) | TRUE |
| MS/MS ammonia loss (FTMS) | TRUE |
| MS/MS dependent losses (FTMS) | TRUE |
| MS/MS recalibration (FTMS) | FALSE |
| MS/MS tol. (ITMS) | 0.5 Da |
| Top MS/MS peaks per Da interval. (ITMS) | 8 |
| Da interval. (ITMS) | 100 |
| MS/MS deisotoping (ITMS) | FALSE |
| MS/MS deisotoping tolerance (ITMS) | 0.15 |
| MS/MS deisotoping tolerance unit (ITMS) | Da |
| MS/MS higher charges (ITMS) | TRUE |
| MS/MS water loss (ITMS) | TRUE |
| MS/MS ammonia loss (ITMS) | TRUE |
| MS/MS dependent losses (ITMS) | TRUE |
| MS/MS recalibration (ITMS) | FALSE |
| MS/MS tol. (TOF) | 40 ppm |
| Top MS/MS peaks per Da interval. (TOF) | 10 |
| Da interval. (TOF) | 100 |
| MS/MS deisotoping (TOF) | TRUE |
| MS/MS deisotoping tolerance (TOF) | 0.01 |
| MS/MS deisotoping tolerance unit (TOF) | Da |
| MS/MS higher charges (TOF) | TRUE |
| MS/MS water loss (TOF) | TRUE |
| MS/MS ammonia loss (TOF) | TRUE |
| MS/MS dependent losses (TOF) | TRUE |
| MS/MS recalibration (TOF) | FALSE |
| MS/MS tol. (Unknown) | 20 ppm |
| Top MS/MS peaks per Da interval. (Unknown) | 12 |
| Da interval. (Unknown) | 100 |
| MS/MS deisotoping (Unknown) | TRUE |
| MS/MS deisotoping tolerance (Unknown) | 7 |
| MS/MS deisotoping tolerance unit (Unknown) | ppm |
| MS/MS higher charges (Unknown) | TRUE |
| MS/MS water loss (Unknown) | TRUE |
| MS/MS ammonia loss (Unknown) | TRUE |
| MS/MS dependent losses (Unknown) | TRUE |
| MS/MS recalibration (Unknown) | FALSE |
| Site tables | Deamidation (NQ)Sites.txt;Oxidation (M)Sites.txt;Phospho (ST)Sites.txt |

2.9) Data analysis and Visualization.

2.9.1) Manually verify the MS/MS spectrum and phosphorylation localization score within PD2.4

2.9.2) Now export the filtered Phosphosites from modifications table for each of the sample/category

2.9.3) Use the below scripts for parsing and combining the data to generate a heatmap representation

*(Note: The below script can also be accessed from the Alessi lab gihub web page: https://github.com/Alessi-Lab/LRRK1\_phosphosites)*

The script below would first read phosphosite mapping result, then map them on to the original protein amino acid sequence through combining PeptideGroups and ModificationSites result text file. The data would be filtered by probability greater or equal to 75 and grouped by the different tryptic digestion enzymes used. Only entries with the highest abundance values according to the unique motif, position and sample condition are kept. Then based on the sequence length, the data was divided into instances of 500 amino acid continuous span on the protein sequence. Each of these instances would be used to create a heatmap where the abundance of the peptide would be the heatmap color, the sample condition would be presented on the X-axis while the position of the phosphosites are represented in the Y-axis in ascending order.

import numpy as np  
import pandas as pd  
from glob import glob  
import re  
import seaborn as sns  
import matplotlib.pylab as plt  
if \_\_name\_\_ == "\_\_main\_\_":  
 proteases = ["AspN", "Chymotrypsin",  
 #"Trypsin"  
 ]  
 files = ["PeptideGroups", "ModificationSites"]  
 phospho\_re = re.compile(r"Phospho [S(\d+)\((\d+)\)]")  
 results = {}  
 for i in glob(r"\\mrc-smb.lifesci.dundee.ac.uk\mrc-group-folder\ALESSI\Toan\TS22D4\_Phosphosite mapping\_02\\*.txt"):  
 for p in proteases:  
 if p in i:  
 for f in files:  
 if f in i:  
 if p not in results:  
 results[p] = {}  
 results[p][f] = pd.read\_csv(i, sep="\t")  
 break  
 break  
 merged\_df = []  
 columns = set()  
 for p in proteases:  
 pg = results[p][files[0]]  
 ms = results[p][files[1]]  
 for i, r in pg.iterrows():  
 pg.at[i, "Primary IDs"] = ";".join([r["Master Protein Accessions"], r["Annotated Sequence"][4:len(r["Annotated Sequence"])-4]])  
 phos = []  
 s = re.search("\[(\d+)-(\d+)\]", r["Positions in Master Proteins"])  
  
 pos = []  
 if s:  
 pg.at[i, "Start"] = s.group(1)  
 mod\_count = r["Modifications"].count("]; ")  
 if mod\_count > 0:  
 for m in r["Modifications"].split("]; "):  
 if "Phospho" in m:  
 s = re.search("\[(.+)", m)  
 if s:  
 for si in s.group(1).split("; "):  
 sire = re.search("(\w)(\d+)\(", si)  
 if sire:  
 phos.append("".join([sire.group(1), sire.group(2)]))  
 pos.append(str(int(sire.group(2)) + int(pg.at[i, "Start"]) - 1))  
 else:  
 if "Phospho" in r["Modifications"]:  
 s = re.search("\[(.+)", r["Modifications"])  
 if s:  
 for si in s.group(1).split("; "):  
 sire = re.search("(\w)(\d+)\(", si)  
 if sire:  
 phos.append("".join([sire.group(1), sire.group(2)]))  
 pos.append(str(int(sire.group(2)) + int(pg.at[i, "Start"]) - 1))  
 pg.at[i, "Position"] = pos  
 pg.at[i, "Phospho"] = phos  
  
 pg = pg.explode(["Phospho", "Position"])  
 pg = pg[pd.notnull(pg["Phospho"])]  
 pg["Position"] = pg["Position"].astype(int)  
 for i, r in ms.iterrows():  
 ms.at[i, "Primary IDs"] = ";".join([r["Protein Accession"], r["Peptide Sequence"]])  
 rpg = pg[[i for i in pg.columns if i.startswith("Abundance")] + ["Primary IDs", "Phospho", "Position", "Modifications"]]  
 rename = {}  
 for i in rpg.columns:  
 if "Abundance" in i:  
 rename[i] = re.sub("Abundance: F\d+: Sample, ", "", i)  
 columns.add(rename[i])  
 print(rpg["Primary IDs"])  
 print(ms["Primary IDs"])  
 rpg = rpg.rename(columns=rename)  
 ms["Phospho"] = ms["Target Amino Acid"] + ms["Position in Peptide"].astype(str)  
 ms["Enzymes"] = p  
 df = ms.merge(rpg, left\_on=["Primary IDs", "Phospho"], right\_on=["Primary IDs", "Phospho"])  
 merged\_df.append(df)  
  
 merged\_df = pd.concat(merged\_df, ignore\_index=True)  
 merged\_df = merged\_df[merged\_df["Site Probability"]>=75]  
 result = pd.melt(merged\_df, id\_vars=[  
 "Phospho", "Position\_y", "Enzymes", "Motif"], value\_vars=list(columns),  
 var\_name="Samples", value\_name="Abundance")  
  
 a = result.groupby([  
 #"Phospho",  
 "Position\_y", "Samples", "Enzymes", "Motif"]).max()  
  
 a.reset\_index(inplace=True)  
 print(a["Samples"])  
 a["Conditions"], a["Replicates"] = a["Samples"].str.split("Rep-", expand=True)  
 for i, g in a.groupby([  
 # "Phospho",  
 "Position\_y", "Motif"]):  
 remove\_motif = True  
 for i2, g2 in g.groupby(["Enzymes", "Conditions"]):  
 if len(g2[pd.notnull(g2["Abundance"])].index) > 1:  
 remove\_motif = False  
 break  
 if remove\_motif:  
 a["Motif"].loc[g.index] = ""  
  
 a.sort\_values("Position\_y", inplace=True)  
 e = 1  
 n = 500  
 samples = a["Samples"].unique()  
 samples\_columns = []  
 for p in proteases:  
 for s in samples:  
 samples\_columns.append((p, s))  
 multiindex = pd.MultiIndex.from\_tuples(samples\_columns, names=["Enzymes", "Samples"])  
 while n:  
  
 c = a[(a["Position\_y"] <= n)&(a["Position\_y"] > (n-500))]  
 fontsize\_pt = plt.rcParams['ytick.labelsize']  
 dpi = 72.27  
 top\_margin = 0.2  
 bottom\_margin = 0.2  
 left\_margin = 0.2  
 right\_margin = 0.2  
 figure\_height = (len(c.index)/10) / (1 - top\_margin - bottom\_margin)  
 figure\_width = 10 / (1-left\_margin-right\_margin)  
 c = c.set\_index([  
 #"Phospho",  
 "Position\_y", "Samples", "Enzymes", "Motif"])  
 c = c.unstack("Enzymes")  
  
 b = pd.pivot\_table(c, values="Abundance", columns="Samples", index=["Position\_y",  
 #"Phospho",  
 "Motif"])  
 b.fillna(0, inplace=True)  
 b = b.T  
  
 for i in b.columns:  
 b0 = b[i][b[i]==0]  
 b[i] = (np.log2(b[i], where=b[i]>0) - np.log2(b[i], where=b[i]>0).mean()) / np.log2(b[i], where=b[i]>0).std(ddof=1)  
 for ind in b0.index:  
 b[i].loc[ind] = np.nan  
 b = b.T  
 new\_df = pd.DataFrame(index=b.index, columns=multiindex)  
 for i in new\_df.columns:  
 if i in b.columns:  
 new\_df[i] = b[i]  
 else:  
 new\_df[i].fillna(0, inplace=True)  
  
 new\_df.to\_csv(f"merged{n}.csv")  
 fig, ax = plt.subplots(  
 figsize=(figure\_width, figure\_height),  
 gridspec\_kw=dict(top=1-top\_margin, bottom=bottom\_margin, left=left\_margin, right=1-right\_margin)  
 )  
 mask = np.isnan(b)  
 sns.heatmap(new\_df, cmap="YlGnBu", mask=mask, square=True, ax=ax)  
 ax.set\_facecolor("silver")  
 ax.xaxis.tick\_top()  
 ax.xaxis.set\_label\_position('top')  
 for label in ax.get\_yticklabels():  
 label.set\_weight("bold")  
 for label in ax.get\_xticklabels():  
 label.set\_weight("bold")  
 plt.xticks(rotation=90)  
 plt.savefig(f"result{n}.pdf")  
 for i, r in b.iterrows():  
 if i[1] != "":  
 p = re.compile(r"[RK]\w[ts]\w\w[RK]")  
 s = re.search(p, i[1])  
 if s:  
 print(i)  
 n += 500  
 e += 1  
 if n >= a["Position\_y"].max():  
 break