CrossMark

Data Article

Proteomic data on the nuclear interactome of human MCM9

James R.A. Hutchins [a,*], Sabine Traver [a], Philippe Coulombe [a], Isabelle Peiffer [a], Magali Kitzmann [a], Daniel Latreille [b,1], Marcel Méchali [a,*]

[a] Laboratory of DNA Replication and Genome Dynamics, Institute of Human Genetics, CNRS, 141 rue de la Cardonille, 34396 Montpellier, France
[b] Laboratory of Gene Regulation, Institute of Human Genetics, CNRS, 141 rue de la Cardonille, 34396 Montpellier, France

ARTICLE INFO

ABSTRACT

We present data relating to the interactome of MCM9 from the nuclei of human cells. MCM9 belongs to the AAA+ superfamily, and contains an MCM domain and motifs that may confer DNA helicase activity. MCM9 has been shown to bind MCM8, and has been implicated in DNA replication and homologous recombination. However, the mechanistic basis of MCM9's role in DNA repair is poorly understood, and proteins with which it interacts were hitherto unknown. We performed tandem affinity purification of MCM9 and its interacting proteins from nuclear extracts of human cells, followed by proteomic analysis, thereby generating a set of mass spectrometry data corresponding to the MCM9 interactome [1]. The proteomic data set comprises 29 mass spectrometry RAW files, deposited to the ProteomeXchange Consortium, and freely available from the PRIDE partner repository with the data set identifier PXD000212. A set of 22 interacting proteins identified from the proteomic data was used to create an MCM9-centered interactive network diagram, using the Cytoscape program. These data allow the scientific community to access, mine and explore the human nuclear MCM9 interactome.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

* Corresponding authors at: Laboratory of DNA Replication and Genome Dynamics, Institute of Human Genetics, CNRS, 141 rue de la Cardonille, 34396 Montpellier, France.
E-mail addresses: james.hutchins@igh.cnrs.fr (J.R.A. Hutchins), marcel.mechali@igh.cnrs.fr (M. Méchali).
[1] Present address: Max Planck Institute of Immunobiology and Epigenetics, Stübeweg 51, D-79108 Freiburg, Germany.

Specifications Table

Subject area	*Biology*
More specific sub-ject area	*Proteomics, Molecular Cell Biology*
Type of data	*Mass spectrometry RAW data files from proteomic analysis of protein complexes*
How data were acquired	*MCM9-containing protein complexes were isolated, proteolytically digested, separated by liquid chromatography, and analysed using an LTQ Velos ion-trap mass spectrometer; raw data were searched against the human IPI database using the Sequest algorithm* [1].
Data format	*Mass spectrometry Thermo RAW data files*
Experimental factors	*Expression of tagged proteins in human cultured cells, preparation of nuclear extracts, tandem protein purification*
Experimental features	*Human MCM9 constructs, tagged at the N- or C-termini with FLAG-HA, were expressed in HeLa S3 cells, and tandem affinity purified. Complexes of MCM9 and associated proteins were analyzed by SDS-PAGE and silver staining, fol-lowed by liquid chromatography-mass spectrometry* [1]. *Peptides and proteins were identified by database searching.*
Data source location	*Institute of Human Genetics (IGH), CNRS, 141 rue de la Cardonille, 34396 Montpellier, France*
Data accessibility	*The data are provided in the public PRIDE repository with the dataset identifier PXD000212. The direct URL to access the data is* http://www.ebi.ac.uk/pride/archive/projects/PXD000212

Value of the data

- The data revealed for the first time the nuclear interactome of human MCM9.
- The data allow the research community to investigate the MCM9 interactome.
- The data act as a benchmark reference of MCM9 interactors for future studies.
- Mining these data could produce additional discoveries about MCM9 interactors.
- Analyzing the data could lead to further insights into MCM9 and its processes.

1. Data

1.1. Mass spectrometry RAW data files

We present 29 mass spectrometry RAW data files that correspond to our proteomic analysis of the human nuclear MCM9 interactome. These data files were deposited to the ProteomeXchange Con-sortium via the PRIDE partner repository [2] with the data set identifier PXD000212 (direct access: http://www.ebi.ac.uk/pride/archive/projects/PXD000212). Data files with names 24770.RAW to 24784.RAW correspond to N-terminally tagged MCM9; 24786.RAW to 24801.RAW correspond to C-terminally tagged MCM9.

1.2. Protein interaction network

We present a protein interaction network centered on MCM9, generated using the Cytoscape program (http://www.cytoscape.org; version 3.2.1) [3] (Fig. 1), based on our proteomic data. The diagram comprises MCM9 plus 22 interacting proteins (combined from the N-terminal and C-terminal FLAG-HA tagged purifications). Contaminant or background proteins have been removed.

2. Experimental design, materials and methods

Full and detailed methods are described in our recent paper [1]. Here we present a summary for each of the steps.

2.1. Generation of cell lines expressing FLAG-HA-tagged MCM9

Stable cell lines expressing double epitope (FLAG-HA; FH) tagged MCM9 at the N- or C-termini were generated using the pOZ retroviral vectors. Vector design ensures tight coupling between the expression of tagged MCM9 and the selection marker, the interleukin-2 receptor α chain (IL2Rα) expressed at the cell surface [4]. Generation of transduction-competent retroviruses was performed by transfecting the retroviral constructs into HEK-293 cells already expressing the retroviral structural gene products required to release retroviral particles. For transduction, HeLa S3 cells were incubated with viral particles containing pOZ-FH-MCM9, pOZ-MCM9-FH, or pOZ-FH (control). Transduced HeLa S3 cells were selected using magnetic affinity beads coupled to an anti-IL2Rα antibody. Following
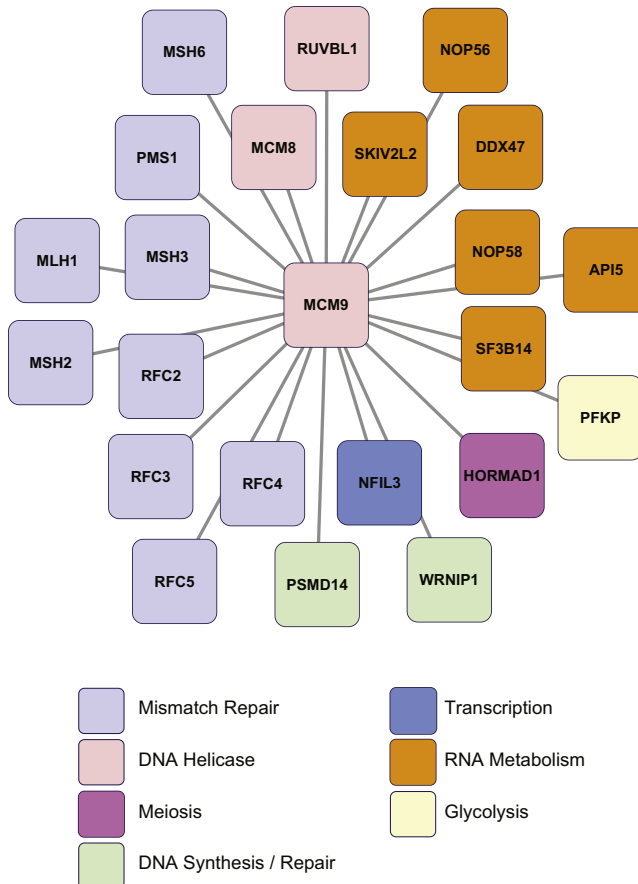

Fig. 1. Nuclear interactome network for human MCM9. The network diagram, generated with Cytoscape software, shows nuclear proteins (nodes), identified by tandem affinity purification and mass spectrometry analysis, that form a complex with human MCM9. Proteins are represented by their official gene symbols, and attributed a color code according to their associated biological processes or functions, as indicated. Note that connecting lines (edges) between proteins do not imply direct physical association, but merely that proteins are members of a shared interactome centered on MCM9. A "live" interactive version of this Cytoscape figure is available on the *Data in Brief* web page for this article.

amplification, cell clones were selected and cultured, and the expression level of double-tagged MCM9 in each clone tested by immunoblotting.

2.2. Tagged protein expression, and tandem affinity purification

For tandem affinity purification, a pool of clones for each construct that showed tagged MCM9 expression levels comparable to those of endogenous MCM9 were used. HeLa S3 cells expressing FLAG-HA-tagged MCM9 at the N- or C-termini, or an empty vector expressing FLAG-HA alone, were grown in exponential culture, and used to prepare soluble extracts of nuclear proteins according to the Dignam method [5]. Isolation of tagged MCM9 and associated proteins was performed using a tandem affinity-purification procedure based on immunoprecipitation of the FLAG and HA tags, according to Nakatani and Ogryzko [4]. In step 1, the nuclear extracts were incubated with anti-FLAG (M2) antibody-conjugated agarose beads. Beads were then washed extensively with FLAG-IP buffer (20 mM Tris–KCl, pH 7.5, 230 mM KCl, 0.03% NP40, 0.07% Tween-20, 1 mM ATP, 5 mM $MgCl_2$), then bound proteins were eluted with two incubations in FLAG peptide. In step 2, the FLAG eluates were incubated with anti-HA antibody-conjugated beads. Beads were washed extensively with HA-IP buffer (20 mM Tris–KCl, pH 7.5, 150 mM KCl, 0.05% NP40, 0.1% Tween-20, 1 mM ATP, 5 mM $MgCl_2$). Bound proteins were eluted with two incubations in HA peptide. MCM9-containing protein complexes were analyzed by SDS-PAGE, and revealed by silver staining.

2.3. Proteomic analysis

The silver-stained gel was sent to the Taplin Biological Mass Spectrometry Facility at Harvard Medical School (http://taplin.med.harvard.edu/) for proteomic analysis. Each SDS-PAGE gel lane was cut into slices, and each processed as a separate sample. Each gel slice was chopped into pieces of approximately 1 mm^3, then in-gel trypsin digested using a modification of a standard procedure [6]. Gel pieces were washed and dehydrated with acetonitrile for 10 min, followed by removal of acetonitrile. Pieces were then completely dried in a SpeedVac centrifugal evaporator (Thermo Scientific). Gel pieces were rehydrated using a solution of 50 mM ammonium bicarbonate (pH 8.0-8.2) containing 12.5 ng/µl modified sequencing-grade trypsin (Promega) at 4 °C. After 45 min, excess trypsin was removed and replaced with 50 mM ammonium bicarbonate, just covering the gel pieces, and samples incubated at 37 °C overnight. Extraction of peptides then proceeded by removing the ammonium bicarbonate solution, followed by one wash with a solution containing 50% acetonitrile, 1% formic acid. Extracts were then dried in a SpeedVac for approximately 1 h, then stored at 4 °C until analysis. For analysis, samples were reconstituted in 5–10 µl HPLC solvent A (2.5% acetonitrile, 0.1% formic acid). A nano-scale reverse-phase HPLC capillary column was prepared by packing 5 µm C18 spherical silica beads (Michrom Bioresources) into a fused silica capillary (125 µm inner diameter $\times$ ~20 cm length) with a flame-drawn tip [7]. After column equilibration, each sample was loaded onto the column using a FAMOS autosampler (LC Packings, Dionex, Thermo Scientific). A gradient was formed and peptides were eluted with increasing concentrations of solvent B (97.5% acetonitrile, 0.1% formic acid). Upon elution, peptides were directly subjected to electrospray ionization and injection into an LTQ Velos ion-trap mass spectrometer (Thermo Scientific). Peptides were detected, isolated and fragmented to produce a tandem mass spectrum of specific fragment ions for each peptide.

2.4. Peptide and protein identification

Identification of peptides and proteins was performed by the data analysis pipeline of the Taplin Facility. MS RAW data files were matched to entries in the human International Protein Index (IPI) database [8] using the Sequest algorithm [9]. All accepted peptides have a cross-correlation (Xcorr) score of at least 0.5. Identified peptides, Xcorr scores and corresponding protein identifiers (IDs) for each sample were presented as HTML tables accessible via a web browser.

2.5. Data post-processing and data-mining

Peptide data were copied from HTML tables and pasted into Microsoft Excel spreadsheets for further processing and analysis. Identified proteins were mapped from their IPI to equivalent Swiss-Prot [10] identifiers using the ID mapping functionality of the UniProt website (http://www.uniprot.org/uploadlists/). Percentage sequence coverage, and numbers of non-distinct and distinct peptides were calculated using the Protein Coverage Summarizer program from the Pan-Omics Research division of the Pacific Northwest National Laboratory (http://omics.pnl.gov/software/protein-coverage-summarizer). UniProt IDs were used to generate a hyperlinked data table, as described [11]. A subset of identified proteins was classified as being "background or contaminant" (at the level of gene symbols) based on two reference sources: the set of "contaminant" proteins published by the MitoCheck consortium [12], and the CRAPome database (http://www.crapome.org/) [13]. Protein entries classified as "background or contaminants" were assigned labels in separate columns in the data table, enabling them to be filtered from the dataset, as required. A table of identified proteins, and a spreadsheet containing all identified proteins and peptides for each sample are available as part of our previous publication [1].

Acknowledgments

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2015.11.055.

References

[1] S. Traver, P. Coulombe, I. Peiffer, J.R.A. Hutchins, M. Kitzmann, D. Latreille, et al., MCM9 is required for mammalian DNA mismatch repair, Mol. Cell 59 (2015) 831–839.
[2] J.A. Vizcaino, R.G. Cote, A. Csordas, J.A. Dianes, A. Fabregat, J.M. Foster, et al., The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013, Nucleic Acids Res. 41 (2013) D1063–D1069.
[3] M.E. Smoot, K. Ono, J. Ruscheinski, P.L. Wang, T. Ideker, Cytoscape 2.8: new features for data integration and network visualization, Bioinformatics 27 (2011) 431–432.
[4] Y. Nakatani, V. Ogryzko, Immunoaffinity purification of mammalian protein complexes, Methods Enzym. 370 (2003) 430–444.
[5] J.D. Dignam, R.M. Lebovitz, R.G. Roeder, Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei, Nucleic Acids Res. 11 (1983) 1475–1489.
[6] A. Shevchenko, M. Wilm, O. Vorm, M. Mann, Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels, Anal. Chem. 68 (1996) 850–858.
[7] J. Peng, S.P. Gygi, Proteomics: the move to mixtures, J. Mass Spectrom. 36 (2001) 1083–1091.
[8] P.J. Kersey, J. Duarte, A. Williams, Y. Karavidopoulou, E. Birney, R. Apweiler, The International Protein Index: an integrated database for proteomics experiments, Proteomics 4 (2004) 1985–1988.
[9] J.K. Eng, A.L. McCormack, J.R. Yates, An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database, J. Am. Soc. Mass Spectrom. 5 (1994) 976–989.
[10] UniProt Consortium, UniProt: a hub for protein information, Nucleic Acids Res. 43 (2015) D204–D212.
[11] J.R.A. Hutchins, What's that gene (or protein)? Online resources for exploring functions of genes, transcripts, and proteins, Mol. Biol. Cell 25 (2014) 1187–1201.

[12] J.R.A. Hutchins, Y. Toyoda, B. Hegemann, I. Poser, J.K. Heriche, M.M. Sykora, et al., Systematic analysis of human protein complexes identifies chromosome segregation proteins, Science 328 (2010) 593–599.

[13] D. Mellacheruvu, Z. Wright, A.L. Couzens, J.P. Lambert, N.A. St-Denis, T. Li, et al., The CRAPome: a contaminant repository for affinity purification–mass spectrometry data, Nat. Methods 10 (2013) 730–736.