**Research data management plan (RDMP)**

|  |
| --- |
| Administrative information |
|  | Principal investigator  | Dr. sc. Iva M. Tolić |
|  | Affiliation | Professor |
|  |  Project proposal title | How centrosome age affects the formation, resolution, and fate of polar chromosomes |
|  | RDMP contact person | Dr. sc. Iva M. Tolić |
| 1. | Data collection and documentation |
|  | What data will you collect, analyse, generate or reuse? (Please state the type, format and volume of data you will collect, not only final data set that will be the result research)  | The project will generate a few major types of raw, primary research data: **i)** microscopy time-lapse movies and images of human cells. These data will be stored in the format they were originally created in, the proprietary software of the microscope company. All these file formats can be read, analysed, and edited using *Image J*, an open-source freely available image analysis software. For presentation purposes, individual images will be stored as TIFF files. The volume of data generated over the course of the project is estimated at 20TB **ii)** Western Blot data will be recorded electronically, stored in the format they were originally created in and converted to TIFF files. The volume of data generated over the course of the project is estimated at 100GB **iii)** Analysis code will be written in *MATLAB* or *R*, and the code will be stored also in *MATLAB* or R format. The volume of data generated over the course of the project is estimated at 1GB. **iv)** Measurements and quantification of the images will be recorded in Excel spreadsheets (estimated volume of 10MB). Analysis of the data will be carried out with *MATLAB* or *R* **v)** Images will be edited with *ImageJ* and mounted into figures with *Adobe Illustrator*. |
|  | How will the data be collected, processed, or generated? (Briefly describe methodologies and quality assurance processes you will use, organization of your project files and data, tools and instruments which will be used for collecting and processing the data)  |  All samples on which data are collected will be prepared according to published standard protocols in the field. All files will be named according to the following convention: **i)** a folder with the name of an experimenter **ii)** a subfolder with the type of methodology (e.g., Western blot, live cell imaging movie etc) **iii)** a sub-subfolder with the type of experiment (e.g., Immunoprecipitation, total cell extract, purified protein) **iv)** a sub-sub-subfolder with the name of the experiment. This name contains the date of the experiment (YYYYMMDD), the cell type ID (will be defined in the README file), the name of the experiment (e.g., PRC1 inhibition), and the name of the main tool (PRC1 antibody). Each folder of a particular experiment will contain a standardized README file giving the name of an experimenter, the date and all the experimental conditions. The individual data file will contain the name of the experiment (e.g., PRC1 inhibition) and the main tool (e.g., PRC1 antibody). Microscopy data are accompanied by metadata files in txt format that indicate the exact recording conditions of an image file. All the quantification files (Excel spreadsheets) will be placed in the folder of the experiment. |
|  | What data documentation and metadata you will develop and provide that are accompanying the data? (In documentation provide all information needed for users to be able to read and interpret the data in the future e. g. code books, ReadMe files, etc.) | Each data set folder of a particular experiment will contain a standardized README file. This README file will include predefined windows details: **i)** the name of an experimenter, **ii)** the date of the experiment- the origin of the sample (cell ID, and all the fluorescent markers or genetic traits e.g. knock-down, knock-in), **iii)** the perturbations applied to the sample (e.g. Depletion of WDR62, inhibition of PRC1, the addition of monastrol), **iv)** the protocol used to prepare the sample (fixation conditions, pre-treatments e.g. cold-treatment), **v)** the tools used to probe the sample (antibody and dilution, PRC oligonucleotide etc.), **vi)** the conditions of recording (exposure times, objectives, temperatures, wavelengths and fluorescence filters, the intensity of the light or laser, for images number of Z-stacks and distance between the stacks, for movies temporal sampling rate, duration of the experiment). This information will allow the other group members to fully understand the origin and generation of the research data. All the microscope images also have a corresponding txt file with a machine-readable metadata file containing the conditions of recording. |
| 2. | Ethics, legal and security issues |
|  | Are you restricted by a confidentiality agreement? Do you have the necessary permission to obtain process, preserve and share the data? Have the people whose data is being preserved been informed or did they give their consent? What methods will you use to ensure the protection of sensitive data (GDPR special category personal data, specify methods of data anonymization)? | There is no restriction by confidentiality agreement. No personal or sensitive data is generated, and therefore no special security steps will be undertaken. However, all computers in the laboratory, as well as access to the NAS storage devices, are password-protected. Access to the data is limited to rights holders (central authentication). |
|  | How will you regulate access to the data and their security? What potential risks do you have to take in consideration? How will you ensure safe sensitive data storage? | Data will be processed and managed in a secure offline environment using virtual desktop technology. |
|  | How will you manage copyright and Intellectual Property Rights issues? Who will be the owner of the data? Which licenses will be applied to the data? What restrictions apply to the reuse of third-party data? | The research result is not expected to lead to a patent. Other intellectual property problems will be solved according to the recommendations of the Ruđer Bošković Institute and University of Zagreb. Since the data is not the subject of a contract it is not anticipated that this study will generate any patentable data or proprietary data which would have to be protected, and it will be released as open data under Creative Commons CC-BY-NC-ND 4.0. |
| 3. | Data storage and preservation |
|  | How will you store different versions of data during the project?How will your data be backed-up during the project?What amount of data are you expecting to be collected and stored during the project (specify in MB/GB/TB)  | All primary (raw) and secondary (raw data analyses) data will be stored on NAS network storage owned by the group. Network-attached storage is typically multiple devices connected that allow data to be stored over high-speed local area network (LAN) connections in a central location that is easily accessible through network protocols and tools. We currently have a space of 60 TB of data. All the secondary data are also located on the laptops of my collaborators, which are themselves backed-up with external hard-disks. |
|  | How will your dataset be curated and preserved during the project and after the project?What file formats will be used for data storage? What amount of data are you expecting to be collected and stored after the project (specify in MB/GB/TB) | The final version of the data set will be shared by the project manager through the institutional repository of the University of Zagreb established in the national Dabar system, where publications and other project documentation will be stored. The data will be released under the Creative Commons CC-BY-NC-ND 4.0. license. We chose the institutional repository in the Dabar system because it supports the principles of FAIR: it assigns the collections a permanent identifier URN:NBN, ensures data visibility through the OpenAIRE and Google Scholar portals and the search engine dabar.srce.hr, and at the same time contributes to the visibility and transparency of the work of the University of Zagreb . |
| 4. | Data sharing and reuse |
|  | How and where will the data be shared? On which repository do you plan to share your data? How will potential users find out about your data? | The data generated will be published in international, open-access peer-reviewed journals. All data that are linked to publications will be accessible at the Dabar and FULIR Full-text Institutional Repository of the Ruđer Bošković Institute, except for the very large live-cell imaging data sets, which we will make available to the community by sending large external hard-disk upon request. The data will be made public at the time of publication. They will be accompanied by the metadata file README.  |
|  | If there is any data which cannot be shared (due to legal, ethical, copyright, confidentiality reasons) explain the reasons of restrictions | No |
|  | Confirm that the digital repository you choose is in line with the FAIR principles | Yes |
|  | Please confirm that you will use a digital repository maintained by a non-profit organisation (if not please explain why) | Yes |