Plan Overview

A Data Management Plan created using DMPonline

Title: How are promoter interactions of cardiac genes differentially rewired in Noonan syndrome with and without a myocardial phenotype?

Creator: Mun-Kit Choy

Principal Investigator: Mun-Kit Choy

Data Manager: Mun-Kit Choy

Project Administrator: Mun-Kit Choy

Affiliation: University of Manchester

Funder: Medical Research Council (MRC)

Template: MRC Template Customised By: University of Manchester

ORCID iD: 0000-0003-3998-0774

Project abstract:

Long-range chromosomal interactions that bring distal regulatory elements such as enhancers into close proximity with gene promoters, play key roles in gene regulation. With BHF support, we previously delineated the first promoter interactome of human embryonic stem cell-derived cardiomyocytes (hESC-CM) using promoter capture HiC technology. We demonstrated that promoter interactions are a key mechanism by which enhancers contact their target genes during hESC-CM differentiation. Specifically, the promoter interactome of hESC-CMs is associated with expression quantitative trait loci (eQTLs) of genes expressed in the left ventricle, and with genome-wide association study (GWAS) regions associated with the heart rate. We now propose to investigate the promoter interactome of cardiomyocytes in a disease context. Noonan syndrome (NS) is a congenital monogenic disorder with a pleomorphic manifestation, characteristically including the cardiovascular system. Hypertrophic cardiomyopathy (HCM) frequently occurs with some causative genes (e.g. RAF1, LZTR1), and infrequently with others (e.g. PTPN11). We will elucidate the perturbation of the promoter interactome during cardiomyocyte development caused by NS causative variants in RAF1, LZTR1 and PTPN11. The findings of this proposed work in single-gene disease will lay a foundation to study promoter interactions of complex congenital heart defects and to identify novel therapeutic targets to mitigate them.

ID: 99375

Start date: 01-01-2023

End date: 31-12-2025

Last modified: 09-11-2023

Copyright information:

How are promoter interactions of cardiac genes differentially rewired in Noonan syndrome with and without a myocardial phenotype?

Manchester Data Management Outline

1. Will this project be reviewed by any of the following bodies (please select all that apply)?

- Funder

2. Is The University of Manchester collaborating with other institutions on this project?

- Yes - Part of a collaboration and owning or handling data

Dr. Mun-Kit Choy is applying for a MRC NIRG. The proposed project will be conducted in collaborations with:

Professor Bernard Keavney (The University of Manchester) will contribute his own time and staff time to oversee the use of the analytical pipelines and statistical models established in Keavney Lab.

Dr. Mikhail Spivakov (Imperial College) will provide advice and consultation for PCHiC, high throughput sequencing and the processing of the raw PCHiC data including the statistical pipeline to identify significant and differential promoter interactions.

Dr. Lukas Cyganek (University of Göttingen) will provide iPSCs derived from NS patients (SFB1002 Research Data Platform; https://sfb1002.med.uni-goettingen.de/) through collaboration and material transfer agreements. Dr. Cyganek will also oversee the maintenance and differentiation of these iPSCs using his laboratory’s optimised protocols.

3. What data will you use in this project (please select all that apply)?

- Acquire new data
- Re-use existing data (please list below)

Quantitative data: Gene expression quantifications (transcriptional [QPCR and transcriptomic sequencing] and translational levels), cellular and phenotyping measurements.

Qualitative data: promoter interactions (genomic positions), whole-genome sequences (raw FASTQ, mapped BAM), chromatograms, images from electrophoresis, immunoblotting, cell imaging and immunohistochemistry.

Existing data: Promoter interactions (genomic positions) in cardiomyocytes derived from human embryonic stem cells published in 2018.

4. Where will the data be stored and backed-up during the project lifetime?

- University of Manchester Research Data Storage Service (Isilon)

5. If you will be using Research Data Storage, how much storage will you require?

- 1 - 8 TB

6. Are you going to be receiving data from, or sharing data with an external third party?

- Yes

Sequencing data will be generated by commercial services offered by companies accredited to ISO/IEC 17025 laboratory standard. The data will be shared with named collaborators' teams in the proposal

7. How long do you intend to keep your data for after the end of your project (in years)?

- 21+ years

Guidance for questions 8 to 13

Highly restricted information defined in the Information security classification, ownership and secure information handling SOP is information that requires enhanced security as unauthorised disclosure could cause significant harm to individuals or to the University and its ambitions in respect of its purpose, vision and values. This could be: information that is subject to export controls; valuable intellectual property; security sensitive material or research in key industrial fields at particular risk of being targeted by foreign states. See more examples of highly restricted information .

Personal information, also known as personal data, relates to identifiable living individuals. Personal data is classed as special category personal data if it includes any of the following types of information about an identifiable living individual: racial or ethnic origin; political opinions; religious or similar philosophical beliefs; trade union membership; genetic data; biometric data; health data; sexual life; sexual orientation.

Please note that in line with data protection law (the UK General Data Protection Regulation and Data Protection Act 2018), personal information should only be stored in an identifiable form for as long as is necessary for the project; it should be pseudonymised (partially de-identified) and/or anonymised (completely de—identified) as soon as practically possible. You must obtain the appropriate ethical approval in order to use identifiable personal data.

8. What type of information will you be processing (please select all that apply)?

- Anonymised personal data

Anonymised genomic data and diagnostic information. DNA sequences have the potential to identify

individuals.

9. How do you plan to store, protect and ensure confidentiality of any highly restricted data or personal data (please select all that apply)?

- Store data in buildings, rooms or filing cabinets with controlled access
- Store data on University of Manchester approved and securely backed up servers or computers
- Anonymise data

High throughput sequencing raw, mapped and processed data will be stored securely in the cluster computer system of the University of Manchester with a capability to back up regularly. A Data Management Plan is in place for this project at the University of Manchester (RDMP 99375).

All other experimental data will be saved, stored and backed up electronically and securely in university-managed computers, shared drives or cloud facilities such as P Drive/Dropbox. Paper/physical records will be stored securely in The University of Manchester.

Storing patients' information is not part of this project. Our collaborator manages that with a separate capacity, following specific ethnic approval and local guidance/rules.

10. If you are storing personal information (including contact details) will you need to keep it beyond the end of the project?

- Not applicable

11. Will the participants' information (personal and/or sensitive) be shared with or accessed by anyone outside of the University of Manchester?

- Not applicable

12. If you will be sharing personal information outside of the University of Manchester, will the individual or organisation you are sharing with be outside the EEA?

- Not applicable

13. Are you planning to use the personal information for future purposes such as research?

- No

14. Will this project use innovative technologies to collect or process data?

- Yes, and innovative technologies will not collect or process personal data (please list the innovative technologies below)

We will use innovative technologies to analyse genomic data for research. Personal data will not be included.

15. Who will act as the data custodian for this study, and so be responsible for the information involved?

munkit.choy@manchester.ac.uk

16. Please provide the date on which this plan was last reviewed (dd/mm/yyyy).

2022-04-30

0. Proposal name

0. Enter the proposal name

How are promoter interactions of cardiac genes differentially rewired in Noonan syndrome with and without a myocardial phenotype?

1. Description of Data.

1.1 Type of Study

High throughput sequencing study: To identify the promoter-genome interactions (promoter interactome) in cardiomyocytes differentiated from induced pluripotent stem cells (iPSCs) derived from Noonan Syndrome (NS) patients caused by known genetic variants with or without myocardial problems, using promoter capture HiC (PCHiC) technology. Transcriptomic sequencing will be performed on these iPSC lines for investigating the effect of promoter interactomic changes on gene expression. Cell lines with the known NS variants repaired will be used as the control lines for all the sequencing experiments.

1.2 Types of Data

Quantitative data: Gene expression quantifications (transcriptional [QPCR and transcriptomic sequencing] and translational levels), cellular and phenotyping measurements.

Qualitative data: promoter interactions (genomic positions), whole-genome sequences (raw FASTQ, mapped BAM), chromatograms, images from electrophoresis, immunoblotting, cell imaging and

immunohistochemistry.
Existing data: Promoter interactions (genomic positions) in cardiomyocytes derived from human embryonic stem cells published in 2018.

1.3 Format and scale of the data

Data will be in .fastq, .bam, .txt, .bed, .washu, .ab1, .jpeg, .tiff, .xls/x, .doc/x, .csv, .pdf, .ppt formats.
Scale: Four NS and four repaired/rescued cell lines from one of the collaborators, Dr Lukas Cyganek (University of Göttingen) will be involved. PCHiC libraries will be prepared from the eight cell lines for PCHiC sequencing. Total RNA will also be extracted from the cell lines for transcriptomic sequencing. Three promoter interaction from PCHiC will be validated in cellular models. Three cell lines with deleted PIRs (promoter-interacting regions) will be generated for each promoter interactions and compared to control cell lines. The data created from these experiments will be stored (and back-up electronically) and analysed. The formats and software used in this research do enable sharing and long-term validity of data.

2. Data collection / generation

2.1 Methodologies for data collection / generation

High throughput sequencing (PCHiC and transcriptomic sequencings) will be performed and data generated by academic or commercial facilities such as Novogene in the UK. Data will be processed and analysed in the Division of Cardiovascular Sciences (The University of Manchester) by the PI in a collaboration with Professor Bernard Keavney and Dr. Mikhail Spivakov. Cellular validation (DNA electrophoresis [images], immunoblotting [images and measurements], gene expression [measurements], cellular and phenotyping measurements) will be carried out in the Division or collaborators' laboratories using established protocols. Chromatograms will be generated from Sanger Sequencing at the Genomic Technologies Core Facility of University of Manchester. Images for immunohistochemistry will be generated using a bright field or fluorescent microscope, digital camera and imaging software. All academic data will be collected to publishable qualities.

2.2 Data quality and standards

We aim to find promoter-genome interactions (and gene expression) in cardiomyocytes differentiated from NS iPSCs. This primary outcome involves high throughput sequencing which will be performed with Illumina platform having ~99% accuracy in sequencing. Each library will be sequenced at the depth as described in Case for Support and mapped uniquely to the reference genome/transcriptome. Mapped data will be quality controlled for depth/coverage of the intended genomic regions. Specific statistical models will be used to filter false positive/background noise. A recommended significance score will be used during the calling of promoter interactions or gene expression.

3. Data management, documentation and curation

3.1 Managing, storing and curating data

High throughput sequencing raw, mapped and processed data will be stored securely in the cluster computer system of the University of Manchester, Research Data Storage (RDS), with a capability to back up regularly. A Data Management Plan is in place for this project at the University of Manchester (RDMP 99375).

All other experimental data will be saved, stored and backed up electronically and securely in university-managed computers, shared drives or cloud facilities such as P Drive/Dropbox. Paper/physical records will be stored securely in The University of Manchester.

Data and files will be named and organised in clear, consistent file structures.

Storing patients' information is not part of this project. Our collaborator in Germany manages that with a separate capacity, following specific ethnic approval and local guidance/rules.

3.2 Metadata standards and data documentation

Data will be reported in a journal together with the necessary methodology. We expect to upload positions of promoter interactions and gene expression data in an appropriate public database/repository, providing documentation or metadata as required by the database/repository in question.

Documentation or metadata will be deposited with the data files enabling users who have no prior knowledge of the research project and the data collected, to understand what the data mean and how the research was carried out so that they can (re)use the data correctly for their respective projects and purposes.

3.3 Data preservation strategy and standards

Electronic data will be held indefinitely, at The University of Manchester's Research Data Storage.

Deposited data such as promoter interactions and gene expression will held at the chosen public repository indefinitely.

Paper/physical records will be archived for 15 years in a secure location.

4. Data security and confidentiality of potentially disclosive personal information

4.1 Formal information/data security standards

The policies and standards to implement and maintain an effective information security management system of participating institutions and ethical committees will be followed. Commercial services needed for the research will be acquired from companies accredited to ISO/IEC 17025 laboratory standard.

4.2 Main risks to data security

The research members in the UK will not have access to patients' information, and only genomic positions and gene expression data will be published with the main outcomes. Therefore the risk of disclosing identifiable patients' information is very low.

Risk management will take place as follows:

- Secure access to data via secure connections e.g. the University's Virtual Private Network (VPN).
- Secure transfer of data e.g. using the University's Dropbox Business service for secure collaboration with externals.
- Physical security of paper records e.g. locked cabinets in secure locations.

5. Data sharing and access

5.1 Suitability for sharing

The data we propose to collect (or existing data we propose to use) in the study is suitable for sharing as it is readily tabulated and digitised.

5.2 Discovery by potential users of the research data

The study protocols, datasets (only genomic position and gene expression data) and results when completed, will be published in relevant journals and deposited in repositories with the indication that data sharing will be appropriate.

Data will be made discoverable through the following steps:

- statement on data availability in publication(s)
- depositing data in well-indexed repositories with rich metadata description
- providing reference on project/institutional website
- publicising information about the data on social media

5.3 Governance of access

Deposited data will be publicly available at the chosen repository such as Gene Expression Omnibus (GEO). For the DNA sequences, participating groups will set up access policies and procedures according to local rules and policies which will be agreed upon the start of the study in the project agreement between institutions. Subjects in this study may be identified through their DNA sequences.

5.4 The study team's exclusive use of the data

This will be decided by the participating groups, with the likely answer being the time period before the publication of the outcomes.

5.5 Restrictions or delays to sharing, with planned actions to limit such restrictions

Translatable outcomes such as drug targets or clinical approaches have potential IP, which would belong to the funder and participating institutions; any delay on sharing will be determined in consultation with the MRC.

5.6 Regulation of responsibilities of users

Deposited data will be publicly available to be used freely. Data sharers of genomic sequences will sign an agreement to have the same responsibilities as members of the study team.

6. Responsibilities

6. Responsibilities

The managers/researchers of this project will have the responsibilities of study-wide data management, metadata creation, data security, and quality assurance of data.

7. Relevant policies

7. Relevant institutional, departmental or study policies on data sharing and data security

The policies that apply to this research are:

The University of Manchester Research Data Management Policy
http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=33802%20

The University of Manchester Records Management Policy
http://documents.manchester.ac.uk/display.aspx?DocID=14916

The University of Manchester Publications Policy
http://documents.manchester.ac.uk/display.aspx?DocID=28526

The University of Manchester IT policies and guidelines
http://www.itservices.manchester.ac.uk/aboutus/policy/

The University of Manchester Intellectual Property Policy
http://documents.manchester.ac.uk/display.aspx?DocID=24420

The University of Manchester Data Protection Policy
http://documents.manchester.ac.uk/display.aspx?DocID=14914

8. Author and contact details

8. Author of this Data Management Plan (Name) and, if different to that of the Principal Investigator, their telephone & email contact details

Dr. Mun-Kit Choy (munkit.choy@manchester.ac.uk; +441613063284; +447881445708)