# The Contribution of Chronic (Retro)Virus Infections to Inflammaging: A Multi-Cohort, Cross-Omics Approach

*A Data Management Plan created using DMPonline.be*

**Creator:** Johan Van Weyenbergh

**Affiliation:** KU Leuven (KUL)

**Funder:** Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

**Template:** FWO DMP (Flemish Standard DMP)

**Principal Investigator:** Johan Van Weyenbergh

**Data Manager:** Johan Van Weyenbergh, n.n. n.n.

**Project Administrator:** Johan Van Weyenbergh

**Grant number / URL:** G0A0621N

**ID:** 200001

**Start date:** 01-01-2021

**End date:** 11-03-2024

**Project abstract:**

The fast aging global population challenges the current public health system. Aging is associated with many changes in the immune system, this project focuses on one of them: chronic low-grade inflammation also known as ‘inflammaging’. We hypothesize that exogenous (HIV-1, HTLV-1) as well as endogenous retroviruses share the ability to induce inflammation through type I interferon (IFN), but that age-, gender- and cell-specific mechanisms define the combined pathogenic effect. If this hypothesis is correct, this logically
implies inflammaging can be objectively quantified and repurposing of anti(retro)viral therapy should be considered in inflammaging. We propose a comprehensive systems biology “cross-omics” approach, to reveal age-specific inflammatory signatures in chronic viral infections (HIV/HTLV-1/CMV). This might allow repurposing of currently used antiviral and antiretroviral drugs for a precision medicine approach, in lifelong treated people living with HIV as well as ‘frail’ elderly coping with inflammaging triggered by herpesviruses (CMV and others).

**Last modified:** 05-07-2023

# The Contribution of Chronic (Retro)Virus Infections to Inflammaging: A Multi-Cohort, Cross-Omics Approach
# FWO DMP (Flemish Standard DMP)

## 1. Research Data Summary

**List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.**

| | | | | Only for digital data | Only for digital data | Only for digital data | Only for physical data |
|---|---|---|---|---|---|---|---|
| Dataset Name | Description | New data | Digital or Physical | Digital Data Type | Digital Data format | Digital data volume (MB/GB/TB) | Physical volume |
| HIV-1 Bahia | 30 PLWH 30 HC TR 60 MB 60 | New data | Digital | Experimental | NA | <1TB | |
| Swiss HIV Cohort | 202 PLWH 6 HC TR 208 | Reuse existing data | Digital | Experimental | NA | <1TB | |
| HTLV-1 Bahia | 20 HAM 20 AC 20 HC TR 60 MB 60 | New data | Digital | Experimental | NA | <1TB | |
| HTLV-1 Sao Paulo | 400 HAM 600 AC GEN 1000 MB 150 | New data | Digital | Experimental | NA | <1TB | |
| HTLV-1 UK | 20 HAM 30 AC 17 HC TR 67 | Reuse existing data | Digital | Experimental | NA | <1TB | |
| BELFRAIL | 549 elderly MB 1098 (0-1y) | New data | Digital | Experimental | NA | <1TB | |
| Vitality 90+ | 146 elderly 30 HC TR 176 | Reuse existing data | Digital | Experimental | NA | <1TB | |
| INCIVAR | 203 elderly 44 HC TR 12 MB 247 | New data | Digital | Experimental | NA | <1TB | |

PLWH: People living with HIV; HAM: HTLV-1-associated myelopathy; AC: asymptomatic HTLV-1 infected controls; HC: healthy controls; TR: transcriptomics; MB: metabolomics; GEN: genomics

**If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:**

| |
|---|
| Swiss HIV Cohort GEO database GSE18233 |
| HTLV-1 UK GEO database GSE29333 |
| Vitality 90+ GEO database GSE40366 |

**Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.**

- No

**Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.**

- No

**Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.**

- No

**Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.**

- No

**Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.**

- No

## 2. Documentation and Metadata

**Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).**

The main body of data of the project consists of transcriptomic data. Since transcriptomics is a rapidly evolving field, we have several technology-specific data sets (micro-array, RNAseq and nCounter). Since they are similar in content, quantification of RNA content in human samples, we are using the same documentation for all data sets, MIAME.
**MIAME** describes the **Minimum Information About a Microarray Experiment** that is needed to enable the interpretation of the results of a experiment unambiguously and potentially to reproduce the experiment. [ Brazma et al. (2001), Nature Genetics]
The six most critical elements contributing towards MIAME are:

1. The raw data for each hybridisation (e.g., CEL or GPR files)
2. The final processed (normalised) data for the set of hybridisations in the experiment (study) (e.g., the gene expression data matrix used to draw the conclusions from the study)
3. The essential sample annotation including experimental factors and their values (e.g., compound and dose in a dose response experiment)
4. The experimental design including sample data relationships (e.g., which raw data file relates to which sample, which hybridisations are technical, which are biological replicates)
5. Sufficient annotation of the array (e.g., gene identifiers, genomic coordinates, probe oligonucleotide sequences or reference commercial array catalog number)
6. The essential laboratory and data processing protocols (e.g., what normalisation method has been used to obtain the final processed data)

1.

**Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.**

- Yes

The main body of data of the project consists of transcriptomic data. SInce transcriptomics is a rapidly evolving field, we have several technology-specific data sets (micro-array, RNAseq and nCounter), which are similar in content, i.e. quantification of RNA content in human samples, we are using the same guidelines for all. MIAME was originally formulated for microarray analysis, but all annotation steps below are similar for RNAseq and nCounter technologies
**MIAME** describes the **Minimum Information About a Microarray Experiment** that is needed to enable the interpretation of the results of a experiment unambiguously and potentially to reproduce the experiment. [ Brazma et al. (2001), Nature Genetics]
The six most critical elements contributing towards MIAME are:

1. The raw data for each experiment (e.g., CEL or GPR files, fastq files for RNAseq, RCC files for nCOunter)
2. The final processed (normalised) data for the set of hybridisations in the experiment (study) (e.g., the gene expression data matrix used to draw the conclusions from the study)
3. The essential sample annotation including experimental factors and their values (e.g., compound and dose in a dose response experiment)
4. The experimental design including sample data relationships (e.g., which raw data file relates to which sample, which hybridisations are technical, which are biological replicates)
5. Sufficient annotation of the array (e.g., gene identifiers, genomic coordinates, probe oligonucleotide sequences or reference commercial array catalog number)
6. The essential laboratory and data processing protocols (e.g., what normalisation method has been used to obtain the final processed data)

## 3. Data storage & back-up during the research project

**Where will the data be stored?**

The data will be stored on our research unit central storage facility (ZIDU server, responsible Prof. Piet Maes). with automatic daily back-up procedures.

**How will the data be backed up?**

Our research unit central storage facility (ZIDU server, responsible Prof. Piet Maes) has automatic daily back-up procedures.

**Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely.**
**If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.**

- Yes

The ZIDU server is constantly updated and storage capacity can be extended if necessary.

**How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?**

The data are stored on our internal ZIDU server and can only be accessed by authorized team members.

**What are the expected costs for data storage and backup during the research project? How will these costs be covered?**

These costs are minimal since this is an internal server for our research unit, no fees need to be paid. Maintenance is provided by our collaborator Prof. Piet Maes.

## 4. Data preservation after the end of the research project

**Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).**

All transcriptomic data from the project will be retained for the expected 5 year period after the end of the project.

**Where will these data be archived (stored and curated for the long-term)?**

Following publication, all transcriptomic data will be archived, as required in the field, into publicly available databases (e.g. GEO, Gene Expression Omnibus), as we have done in the past with our published transcriptomic data.

**What are the expected costs for data preservation during the expected retention period? How will these costs be covered?**

The costs are minimal since this is an internal server for our research unit, no fees need to be paid. Maintenance is provided by our collaborator Prof. Piet Maes.

## 5. Data sharing and reuse

**Will the data (or part of the data) be made available for reuse after/during the project?  In the comment section please explain per dataset or data type which data will be made available.**

- Yes, in an Open Access repository

Following publication, all transcriptomic data will be archived, as required in the field, into publicly available databases (e.g. GEO, Gene Expression Omnibus), as we have done in the past with our published transcriptomic data.

**If access is restricted, please specify who will be able to access the data and under what conditions.**

No restrictions to access

**Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.**

- No

**Where will the data be made available? If already known, please provide a repository per dataset or data type.**

Publicly available databases (e.g. GEO, Gene Expression Omnibus), as we have done in the past with our published transcriptomic data.

**When will the data be made available?**

Upon publication of the research results.

**Which data usage licenses are you going to provide? If none, please explain why.**

As is standard practice in the field of transcriptomics, all data (anonymized and without any sensiive personal data) are made publicly available in open databases for the entire scientific community, without restriction.

**Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.**

- Yes

Not available yet, only after publication.

**What are the expected costs for data sharing? How will these costs be covered?**

No cost, open databases such as GEO are free.

## 6. Responsibilities

**Who will manage data documentation and metadata during the research project?**

Johan Van Weyenbergh

**Who will manage data storage and backup during the research project?**

Johan Van Weyenbergh and Piet Maes

**Who will manage data preservation and sharing?**

Johan Van Weyenbergh

**Who will update and implement this DMP?**

Johan Van Weyenbergh

# The Contribution of Chronic (Retro)Virus Infections to Inflammaging: A Multi-Cohort, Cross-Omics Approach
# Application DMP

## Questionnaire

**Describe the datatypes (surveys, sequences, manuscripts, objects … ) the research will collect and/or generate and /or (re)use. (use up to 700 characters)**

Table I: List of cohorts and datasets available

| Cohort | Patients (n) | Matched Healthy Controls (n) | Omics data (n) |
|---|---|---|---|
| Retroviral | | | |
| HIV-1 Bahia | 30 PLWH | 30 HC | TR 60 MB 60 |
| Swiss HIV Cohort | 202 PLWH | 6 HC | TR 208 |
| HTLV-1 Bahia | 20 HAM | 20 AC 20 HC | TR 60 MB 60 |
| HTLV-1 Sao Paulo | 400 HAM | 600 AC | GEN 1000 MB 150 |
| HTLV-1 UK | 20 HAM | 30 AC 17 HC | TR 67 |
| Inflammaging | | | |
| BELFRAIL | 549 | - | MB 1098 (0-1y) |
| INCIVAR | 203 | 44 HC | TR 12 MB 247 |
| Vitality 90+ | 146 | 30 HC | TR 176 |
| | | | |
| Total | 1570 | 897 | TR 583 MB 1615 |

PLWH: People living with HIV; HAM: HTLV-1-associated myelopathy; AC: asymptomatic HTLV-1 infected controls; HC: healthy controls; TR: transcriptomics; MB: metabolomics; GEN: genomics

**Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)**

1. Designation of responsible person: Johan Van Weyenbergh
2. Storage capacity/repository
    - during the research: Where will the data be stored?
      1. The time-stamped master copy of the data will be kept on our research unit central storage facility (ZIDU server, responsible Prof. Piet Maes). Copies can be made and kept on personal devices.
      2. Since we will collaborate with researchers from other research units and groups, we will use OneDrive for active use of the data during the project.
      How is backup of the data provided?
      The data will be stored on our research unit central storage facility (ZIDU server, responsible Prof. Piet Maes). with automatic daily back-up procedures.
      Is there currently sufficient storage & backup capacity during the project? Yes, the ZIDU server is constantly updated and storage capacity can be extended if necessary.
      What are the expected costs for data storage and back up during the project? How will these costs be covered?
      These costs are minimal since this is an internal server for our research unit, no fees need to be paid. Maintenance is provided by our collaborator Prof. Piet Maes.
      Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?
      The data are stored on our internal ZIDU server and can only be accessed by authorized team members.
    - Data preservation after the FWO project
      Which data will be retained for the expected 5 year period after the end of the project?
      All transcriptomic data from the project will be retained for the expected 5 year period after the end of the project.
      Where will the data be archived (= stored for the longer term)?
      Following publication, all transcriptomic data will be archived, as required in the field, into publicly available databases (e.g. GEO, Gene Expression Omnibus), as we have done in the past with our published transcriptomic data.
      What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?
      The costs are minimal since this is an internal server for our research unit, no fees need to be paid. Maintenance is provided by our collaborator Prof. Piet Maes.

**What’s the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)**

NA

**Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)**

No personal or sensitive data are used.

**Which other issues related to the data management are relevant to mention? (use up to 700 characters)**

NA

# The Contribution of Chronic (Retro)Virus Infections to Inflammaging: A Multi-Cohort, Cross-Omics Approach
# DPIA

## DPIA

**Have you performed a DPIA for the personal data processing activities for this project?**

- Not applicable

# The Contribution of Chronic (Retro)Virus Infections to Inflammaging: A Multi-Cohort, Cross-Omics Approach

## GDPR

### GDPR

**Have you registered personal data processing activities for this project?**

- No