

---

## DMP du projet "DUBii2021\_colon\_cancer"

*Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "Science Europe - DMP template (english)" fourni par Science Europe.*

### Plan Details

<b>Plan title</b>	DMP du projet "DUBii2021_colon_cancer"
<b>Language</b>	fra
<b>Creation date</b>	2021-06-12
<b>Last modification date</b>	2021-06-12
<b>Identifiant</b>	

### Project Details

<b>Project title</b>	DUBii2021_colon_cancer
<b>Abstract</b>	<p>The main goal of the project is to study an evolution of immune components in colon-cancer data deconvoluted by Factor-forest application.</p> <p><a href="https://github.com/NicolasSompairac/Hierarchical_ICA">https://github.com/NicolasSompairac/Hierarchical_ICA</a></p>

### Research outputs :

1. Omics data analysis resulting from clinical trials produced in IMMUCAN European project (Jeu de données)

### Contributors

Name	Affiliation	Roles
Maria Kondratova		<ul style="list-style-type: none"><li>• Coordinateur du projet</li><li>• Personne contact pour les données</li><li>• Responsable du plan</li></ul>

### Droits d'auteur :

Le(s) créateur(s) de ce plan accepte(nt) que tout ou partie de texte de ce plan soit réutilisé et personnalisé si nécessaire pour un autre plan. Vous n'avez pas besoin de citer le(s) créateur(s) en tant que source. L'utilisation de toute partie de texte de ce plan n'implique pas que le(s) créateur(s) soutien(nen)t ou aient une quelconque relation avec votre projet ou votre soumission.

# DMP du projet "DUBii2021\_colon\_cancer"

---

## 1. Data description and collection or re-use of existing data

### 1a. How will new data be collected or produced and/or how will existing data be re-used?

Public available data was used for the study.

---

### 1b. What data (for example the kind, formats, and volumes), will be collected or produced?

CRC datasets: TCGA, KFSYSCC, Colonomics, GSE14333, GSE17536, GSE37892, GSE38832, GSE39582, OUS\_125, OUS\_141, Petacc3 were used for primary Factor forest application  
Format: counts, spreadsheets

Single-cell data was downloaded from GEO  
<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE146771>

Format: preprocessed and normalized data, spreadsheets

Proteomic data was downloaded from supplementary of the article:  
"Proteogenomic characterization of human colon and rectal cancer"  
Nature. 2014 Sep 18; 513(7518): 382-387.

Format: preprocessed and normalized data, spreadsheets

---

## 2. Documentation and data quality

### 2a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

Metadata is provided

---

### 2b. What data quality control measures will be used?

No additional quality control, because data is already published and we suppose that quality control has used yet

---

## 3. Storage and backup during the research process

### 3a. How will data and metadata be stored and backed up during the research?

Data does not need back up

---

**3b. How will data security and protection of sensitive data be taken care during the research?**

Used data is public, it does not need a special security protection

---

**4. Legal and ethical requirements, codes of conduct**

**4a. If personal data are processed, how will compliance with legislation on personal data and on security be ensured?**

no

---

**4b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?**

Public data

---

**4c. What ethical issues and codes of conduct are there, and how will they be taken into account?**

no

---

**5. Data sharing and long-term preservation**

**5a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?**

No restriction, data is publicly available, as well as results of the research

---

**5b. How will data for preservation be selected, and where data will be preserved long-term (for example a data repository or archive)?**

Data will be saved on git-hub

---

**5c. What methods or software tools are needed to access and use data?**

No special software

---

**5d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?**

---

## **6. Data management responsibilities and resources**

**6a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?**

Maria Kondratova

---

**6b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?**