



**Identification of functionally active genomic features relevant to phenotypic diversity
and plasticity in cattle**

Deliverable D9.4

Report on training course/hackathon on bioinformatics methods

Grant agreement no°: 815668

Due submission date

2020-08-31

Actual submission date

2021-02-27

Johanna Vilkki

LUKE, johanna.vilkki@luke.fi

Confidential: No

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 815668. The content of this report reflects only the author's view. The European Commission is not responsible for any use that may be made of the information it contains.

DOCUMENT CONTROL SHEET

Deliverable name	Report on training course/hackathon on bioinformatics methods
Deliverable number	D 9.4
Partners providing input to this Deliverable	CRG, LUKE, EAAP
Draft final version circulated by lead party to: On date	FBN on 25 Februar 2021
Approved by (on date)	FBN as Coordinator (2021-02-27)
Work package no	9
Dissemination level	Public (PU)

REVISION HISTORY

Version number	Version date	Document name	Lead partner
1	2021-02-22	D9.4	LUKE
1.1	2021-02-23	D9.4	LUKE
Final	2021-02-26	D9-4-bioinfo-training-final.docx	LUKE

Changes with respect to the DoA (Description of Action)

The timing of the training course was delayed (from August 2020 to November 2020), and the format changed from a physical to an online event, due to the Covid-19 pandemic. The related D9.4 is delivered now in month 18 instead of month 12.

Dissemination and uptake

This deliverable is public and will be available on the BovReg Website. The invited plenary lectures and flash talks were streamed to the BovReg YouTube channel and all of them are still accessible on this playlist

https://www.youtube.com/channel/UC0qhdZc9pDT_6dEk3qDUvcA/playlists.

Each session available on YouTube was reposted and retweeted, one per week, in order to give visibility to the workshop and populate the social media accounts.

Table of Content

1. Summary of results	3
2. Introduction	4
3. Core report	5
Table 1. Participants of the workshop, their position and institute.....	8
Table 2. Participation in the different hackathon sessions each day	9
Hackathon projects	10
Further dissemination of the workshop content	10
4. Conclusions	11
5. References	12
6. Annex 1.....	13

1. Summary of results

A training course/hackathon on bioinformatics methods was organised as an online workshop by CRG from November 17th to 20th 2020, under the name “**Reproducible Genomics Workflows using Nextflow and Nf-core**”.

A total of 41 participants were accepted for participating in the workshop, of these 18 were from the BovReg project, 8 from the GENE-SWitCH project, 2 from the AQUA-FAANG project and 12 from other global FAANG projects, as well as one from an unrelated project. The workshop included a Nextflow tutorial and a hackathon on topics proposed by participants, lectures by keynote speakers, and contributed flash talks selected from participants’ abstracts. The invited plenary lectures and flash talks were streamed to the BovReg YouTube channel where all of them are still available.

2. Introduction

WP9 has the objective to conduct training, communication and efficient knowledge transfer within the BovReg project, including joint activities with the cluster of EU FAANG projects, global FAANG initiatives, and the related industry. One aim is to identify and implement options of achieving synergy across the three cluster projects funded under the SFS-30-2018 call topic. In the training plan, BovReg included three courses, and this deliverable now reports on the second one, which was held in November 2020.

- 1) Hands-on course on laboratory methodology, 5 days, hosted by UNILIM, with possible involvement of ULIEGE, Diagenode (postponed due to Covid-19 restrictions)
- 2) Hands-on course on bioinformatics pipelines and data management, 3 days, hosted by CRG with content based on the feedback of the first survey (D3.1) and agreements at a joint meeting of SFS-30-2018 call funded EU H2020 FAANG projects on bioinformatic pipelines used across cluster projects; and
- 3) Integration of biological information into genomic evaluation models, 4 days, hosted by AU (scheduled for month 42 of the project)

The original plan was to have the hands-on on-site workshop/hackathon about bioinformatics pipelines during summer 2020 for maximum use within BovReg and other EuroFAANG projects. The Covid-19 pandemic, however, hampered travelling and organising of an onsite meeting, and therefore the organising committee started planning an online event. The planning committee consisted of Johanna Vilkki and Daniel Fischer from LUKE, Cedric Notredame, Jose Espinosa-Carrasco and Damjana Kastelic from CRG, Andrea Amaral from FMV, Christa Kuehn from FBN, Daniel Zerbino from EMBL-EBI and Gabriel Costa from ULIEGE. Meetings of the committee were held online on 6th and 21st July 2020.

It was decided, to organise the virtual workshop for the animal genomics community in November 2020 with CRG as the lead party. The focus would be to learn how to implement bioinformatics workflows, following best practices in terms of computational reproducibility and interoperability using Nextflow [1]. Nextflow has become one of the more popular bioinformatics workflow managers. Its large set of features provides the toolkit to implement fully portable, replicable, scalable workflows in a seamless manner.

The workshop was planned to include a Nextflow tutorial and hackathon, lectures by keynote speakers, and contributed flash talks selected from participants' abstracts. The name of the workshop was: **Reproducible Genomics Workflows using Nextflow and NF-Core**, to be held from 17th November to 20th November 2020.

It was also decided that BovReg participants would have the first opportunity to register, but that places would be reserved also for other EuroFAANG projects (AQUA-FAANG and

GENE-SWitCH) as well as global FAANG members. Originally, the participant number was restricted to 30, but due to the high interest, it was expanded to 41 virtual participants.

3. Core report

The workshop in practice was organized by Courses@CRG and the CRG team, including Cedric Notredame, Jose Espinosa-Carrasco, Luca Cozzuto, Leila Mansouri, Edgar Garriga, Athanasios Baltzis and Suzanne Jin.

As keynote speakers the organising committee invited: Paolo Di Tommaso and Evan Floden from Sequera Labs (Spain, formerly at CRG), Luca Cozzuto from CRG (Spain), Phil Ewels from SciLifeLab (Stockholm) and Harshil Patel from The Francis Crick Institute (UK).

The workshop was announced through the BovReg website and social media:

Website: <https://www.bovreg.eu/crg-online-workshop-november-17-20-2020/>

Google Sites: <https://sites.google.com/view/reproduciblegenomicsworkflowsu/home>,

Twitter account: <https://twitter.com/BovReg/status/1311612905444511745>

Facebook account: <https://www.facebook.com/BovReg.H2020/posts/220294312759878>

The full workshop programme is inserted overleaf.

The organizing committee got 84 applications for attendance. To maintain the group of attendees in the training sessions manageable the number of attendees was restricted to 41. All other sessions like talks, flash talks and hackathon were open to the whole community. The hackathon part was more or less self-organized by some of the participants.

The participants were accepted based on their submitted interest letters. The participants comprised a good representation from the three EU H2020 FAANG cluster projects BovReg (18 participants), GENE-SWitCH (8), and AQUA-FAANG (2), as well as global FAANG initiatives (12) and related projects (1) (Table 1).

Of the accepted participants, 28 were male and 13 were female.

Bioinformatic Training Course Programme

<p>TUESDAY 17th November 2020</p> <p>Chair: Cedric Notredame (CRG)</p> <p>9.30 - 9.45 - Welcome - Cedric Notredame Centre for Genomic Regulation (CRG) - YouTube Streaming</p> <p>9.45 - 10.30 - Introduction to Nextflow - Evan Floden Sequera Labs - YouTube Streaming</p> <p>10.30 - 11.20 - Nextflow training* (for accepted participants only)</p> <p>11.20 - 11.30 - <i>Break</i></p> <p>11.30 - 12.30 - Nextflow training* (for accepted participants only)</p> <p>12.30 - 13.00 - Flash talks - YouTube Streaming</p>	
	<ul style="list-style-type: none"> ▪ <i>From Snakemake to Nextflow</i> - Daniel Fischer Natural Resources Institute Finland (Luke) ▪ <i>Dynamics of direct transmission of Carbapenem resistance in Enterobacterales between human families and their companion animal</i> - Juliana Menezes Faculdade de Medicina Veterinária, Universidade de Lisboa *this talk will not be streamed
<p>13.00 - 14.30 - <i>Break</i></p> <p>14.30 - 15.00 - HACKATHON: introduction and presentation of projects - Zoom Session</p> <p>14.30 - 17.30 - HACKATHON - Zoom Sessions</p> <p>WEDNESDAY 18th November 2020</p> <p>Chair: Andreia Amaral (FMV)</p> <p>9.30 - 11.20 - Nextflow training* (for accepted participants only)</p> <p>11.20 - 11.30 - <i>Break</i></p> <p>11.30 - 12.15 - Introduction to nf-core - Phil Ewels SciLifeLab - YouTube Streaming</p> <p>12.15 - 13.00 - Flash talks - YouTube Streaming</p>	
	<ul style="list-style-type: none"> ▪ <i>BOmA - Bovine Omics Atlas</i> - Siddharth Jayaraman The Roslin Institute ▪ <i>Structural variations in the cattle genomes</i> - Bala Kiran Manthri SLU University ▪ <i>Detection of Mobile genetic elements insertion sites in Swiss cattle breeds</i> - Meenu Bhati ETH, Zurich ▪ <i>GWA, mQTL and eQTL studies with imputed whole genome variants for BovReg traits</i> - Praveen Krishna Chitneedi Leibniz Institute for Farm Animal Biology (FBN)
<p>13.00 - 14.30 - <i>Break</i></p> <p>14.30 - 15.00 - Hackathon advances from the previous day - Zoom Session</p> <p>14.30 - 17.30 - HACKATHON - Zoom Sessions</p>	

THURSDAY 19th November 2020

Chair: Daniel Fischer (Luke)

9.30 - 11.20 - Nextflow training* (for accepted participants)

11.20 - 11.30 - Break

11.30 - 12.30 - Featured nf-core pipelines - **Harshil Patel** | [The Francis Crick Institute](#) - [YouTube Streaming](#)

12.30 - 13.00 - Flash talks - [YouTube Streaming](#)

- *The evolution of the IsomiR Window, an upgrade to nextflow and nf-core* - **Andreia Amaral** | [Faculdade de Medicina Veterinária da Universidade de Lisboa](#)
- *Functional and Structural Genomics Annotation* - **Gabriel Moreira** | [University of Liège](#)
- *Large scale sequence level imputation for dairy industry* - **Tuan Nguyen** | [Agriculture Victoria](#)

13.00 - 14.30 - Break

14.30 - 15.30 - Running nf-core pipelines + Contributing to nf-core - **Phil Ewels** | [SciLifeLab](#) - [YouTube Streaming](#)

15.30-15:40 Break

15.40 - 16.10 - Hackathon advances from the previous day - Zoom Session

16.00 - 17.30 - HACKATHON - Zoom Session

FRIDAY 20th November 2020

Chair: Jose Espinosa-Carrasco ([CRG](#))

9.30 - 11.20 - Nextflow training* (for accepted participants)

11.20 - 11.30 - Break

11.30 - 12.30 - Introduction to DSL2 - **Paolo di Tommaso** | [Sequera Labs](#) - [YouTube Streaming](#)

12.30 - 13.00 - Flash talks - [YouTube Streaming](#)

- *Improving genome annotations with RNA-seq data: a new pipeline to combine transcript reconstruction and expression assessment* - **Cyril Kurylo** | [INRAE](#)
- *Link-HD: a flexible tool to integrate and explore association between multiple microbial communities* - **Yuliaxis Ramayo Caldas** | [IRTA](#)
- *GENE-SWitCH methylation pipeline for coherent methylation analysis* - **Jani de Vos** | [Wageningen University & Research](#)

13.00 - 14.30 - Break

14.30 - 15.00 - Master of Pores: "A nanopore processing pipeline" - **Luca Cozzuto** | [CRG](#) - [YouTube Streaming](#)

15.00 - 17.00 - HACKATHON - Zoom Sessions

17.00 - 17.30 - Presentation of the HACKATHON projects and closing - Zoom Session

Table 1. Participants of the workshop, their position and institute

	Position, Project	Institute, Country
1	PI, GENE-SWitCH	INRAE, France
2	Bioinformatician, BovReg	EMBL-EBI, UK
3	MSc Student, BovReg	CRG; Spain
4	Researcher, BovReg	Faculdade de Medicina Veterinária da Universidade de Lisboa (FMV), Portugal
5	Researcher, Global FAANG	SLU University, Sweden
6	Researcher, BovReg	INRAE, France
7	Postdoc, BovReg	CRG, Spain
8	Researcher, Global FAANG	Mohammed V University, Morocco
9	Researcher, Global FAANG	University of Idaho, USA
10	Research engineer, GENE-SWitCH	INRAE, France
11	MSc student, AQUA-FAANG	George August University Göttingen, Germany
12	researcher, GENE-SWitCH	INRAE, France
13	Bioinformatician, BovReg	Natural Resources Institute Finland (LUKE), Finland
14	Research engineer, BovReg	Leibniz Institute for Farm Animal Biology (FBN), Germany
15	Post Doc, BovReg	University of Liège, Belgium
16	PhD Student, GENE-SWitCH	Wageningen University and Research, The Netherlands
17	PhD Student, BovReg	CIISA, Portugal
18	Staff Scientist, GENE-SWitCH	Wageningen University, The Netherlands
19	Associate Researcher, AQUA-FAANG	Santiago de Compostela University, Spain
20	Researcher, BovReg	FMV, Portugal
21	Platform manager, BovReg	INRAE, France
22	Senior Lecturer, Global FAANG	Curtin University, Australia
23	Research Intern, Global FAANG	International Livestock Research Institute, Nigeria
24	Bioinformatician, BovReg	INRAE, France
25	Head of Genetics Unit, Global FAANG	AFBI, UK
26	Research engineer, BovReg	INRAE, France
27	PhD Student, BovReg	ETH, Zurich, Switzerland
28	FAO National Coordinator on Animal Genetic Resources (AnGR), Global FAANG	Animal Research Institute of Iran (ASRI), Iran
29	MSc, BovReg	Faculdade de Medicina Veterinária da Universidade de Lisboa, Portugal
30	Researcher, Global FAANG	SLU, Sweden
31	Post doc, BovReg	Leibniz Institute for Farm Animal Biology (FBN), Germany

32	Researcher, GENE-SWitCH	INSERM, France
33	Researcher, BovReg	The Roslin Institute, UK
34	PI, GENE-SWitCH, Global FAANG	INRAE, France
35	PI, Global FAANG	Sveriges lantbruksuniversitet, Sweden
36	Research Scientist, Global FAANG, BovReg	Agriculture Victoria, Australia
37	Postdoctoral researcher, Global FAANG	North Dakota State University, USA
38	Researcher, Global FAANG	The Center for Research in Agricultural Genomics, Spain
39	Researcher, GENE-SWitCH	IRTA, Spain
40	Student, Global FAANG	UT, IRAN
41	Research associate, Other	University of the Highlands and Islands, UK

Nextflow training was organised each morning for the accepted participants. A total of 12 flash talks to be given by the participants during the workshop were ranked and selected by the organising/planning committee based on the abstracts sent.

The invited plenary lectures and flash talks were open for anyone interested and streamed to the BovReg YouTube channel (Figure 1).

A hackathon was organised by ZOOM sessions each afternoon. It was open to the whole community. The hackathon was organised as specific projects (Table 2).

Table 2. Participation in the different hackathon sessions each day

No	Name	No. of participants at 4pm			
		Day 1	Day 2	Day 3	Day 4
1	<i>From Snakemake to Nextflow</i>	9	5	6	5
2	<i>WGS data analysis pipeline</i>	14	10	8	10
3	eQTLs analysis	2	0	0	0
4	<i>RNA-seq analysis pipeline</i>	7	7	4	4
5	Fitness traits analysis pipeline	0	0	0	0
6	Environmental effect prediction by omics data integration	0	0	0	0
7	<i>Bacterial DNA analysis pipeline</i>	4	3	7	0
8	Individual projects room	0	5	4	3
TOTAL		36	30	29	22

Four of the proposed hackathon projects did reach sufficient interested participants, and these hackathon projects are available on GitHub: <https://github.com/BovReg/nf-workshop20>.

Hackathon projects

Project 1 - From Snakemake to Nextflow

- Slack channel: [#hackathon-nov-2020-pr1-snakemake-nextflow](#)
- Slides on HackMD: [Edit](#), [view](#)
- Link to the fork of the [GitHub project repository](#) on BovReg

Project 2: WGS data analysis pipeline

- Slack channel: [#hackathon-nov-2020-pr2-wgs-pipeline](#)
- Slides on HackMD: [Edit](#), [view](#)
- Link to the fork of the [GitHub project repository](#) on BovReg

Project 4: RNA-seq analysis pipeline

- Slack channel: [#hackathon-nov-2020-pr4-rnaseq-pipeline](#)
- Slides on HackMD: [Edit](#), [view](#)
- Link to the fork of the [GitHub project repository](#) on BovReg

Project 7: Bacterial DNA analysis pipeline

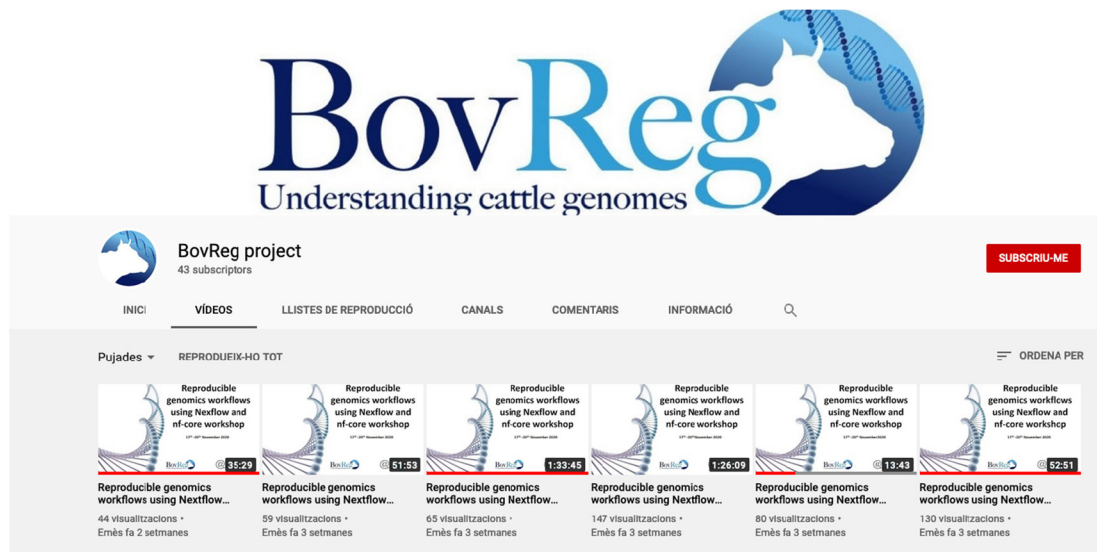
- Slack channel: [#hackathon-nov-2020-pr7-bacterial-dna-pipeline](#)
- Slides on HackMD: [Edit](#), [view](#)
- Link to the fork of the [GitHub project repository](#) on BovReg

Further dissemination of the workshop content

All of the invited lectures and flash talks are still accessible on this playlist: https://www.youtube.com/channel/UC0qhdZc9pDT_6dEk3qDUvcA/playlists (Figure 1). Each session available on YouTube has been reposted and retweeted, one per week in order to give visibility to the workshop and populate the social media accounts. When possible, speakers have been recalled by their nickname on Twitter to let them directly retweet their talk.

Figure 1. The BovReg Youtube channel:

https://www.youtube.com/channel/UC0qhdZc9pDT_6dEk3qDUvcA/playlists



The hackathon projects are still available on GitHub: <https://github.com/BovReg/nf-workshop20>.

4. Conclusions

The workshop gained a lot of interest across Europe (Figure 2) and globally and got very positive feedback from the participants (Figure 3, Annex 1). Very high ratings are given to the Tutorial webpage and the key notes tutorials (e.g. https://bovreg.github.io/nf-workshop20/#_channels), which will provide a legacy to the project and the global FAANG community. The lowest ratings were given for the hackathon part, potentially due to the very diverse and project-specific interests of the participants.

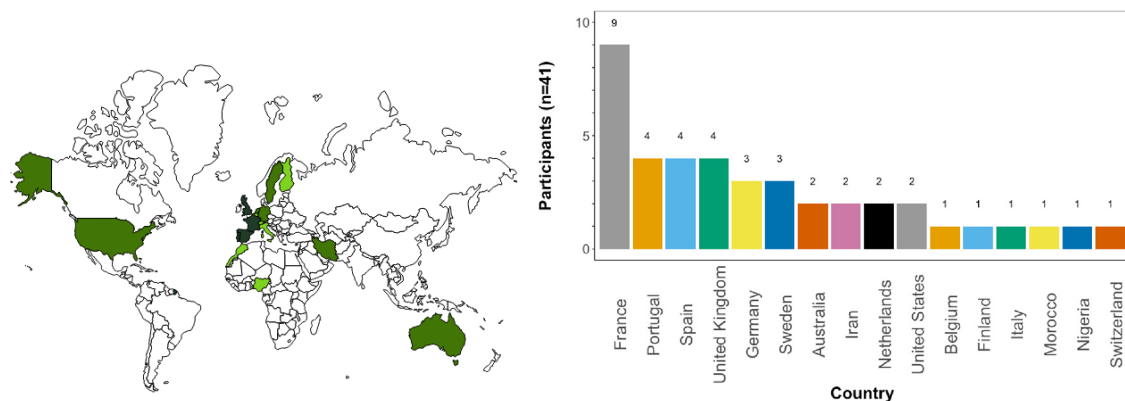


Figure 2. The global distribution of workshop participants across countries

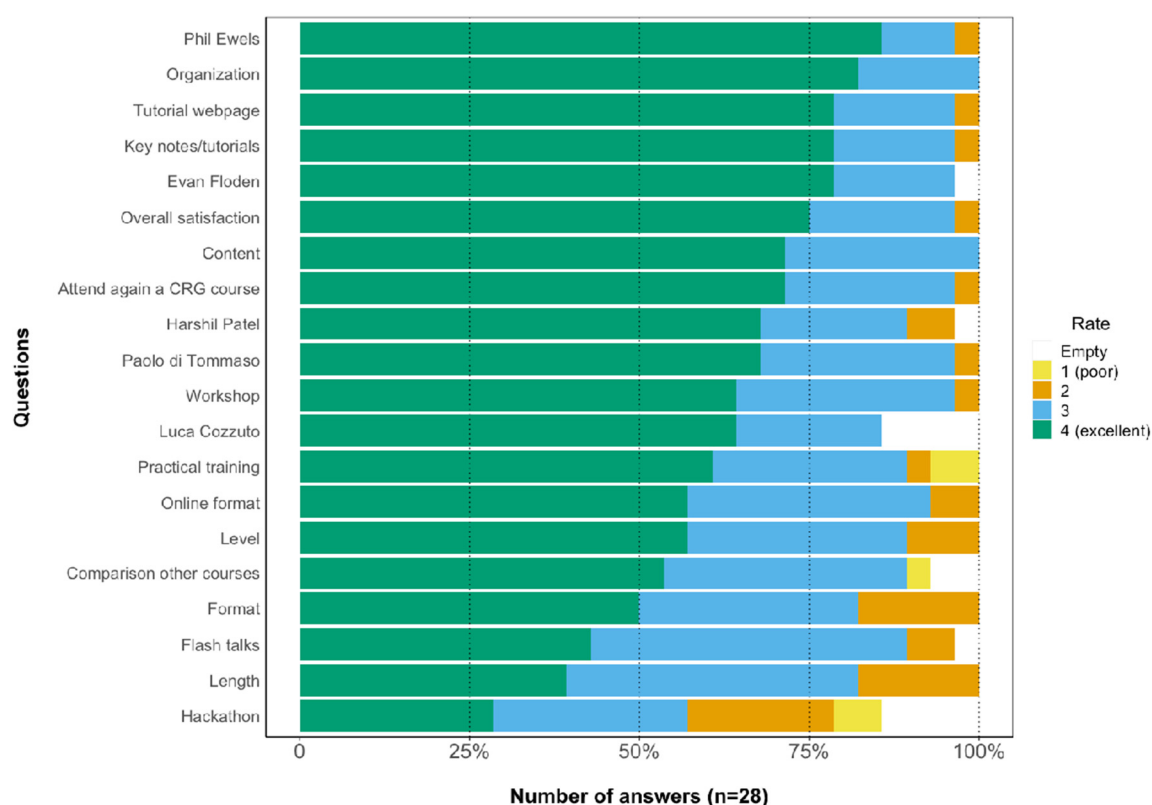


Figure 3. Results from the feedback survey among 28 participants. Most of the categories gained over 50% rating as excellent, with overall satisfaction even at 75%.

5. References

1. Ewels, P.A., Peltzer, A., Fillinger, S. et al. The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol* 38, 276–278 (2020). <https://doi.org/10.1038/s41587-020-0439-x>

6. Annex 1

Some examples from the collected feedback:

Why would you recommend this workshop (or not)?

- It provided a very nice overview and motivated to start using Nextflow
- An excellent overview and step-by-step guide to getting started and presented by team behind Nextflow with expert feedback and solutions to questions.
- Very complete, very well organized, enough time, not too long, not too intensive.
- Very good overview of Nextflow, nf-core
- Would recommend: very informative. Directly applicable, but also providing insights into design philosophy
- It was a new platform and this workshop covered the details of pipelines. Also, it accepted the new ideas as HACKATHON event.
- The workshop is well organized. The vital part of the workshop is the practical experience of the tool. The instructors did a great job helping us.
- It was organized in a way to accommodate all category of linux user
- This workshop was well-structured, insightful, rich in content and provided excellent information.
- I have learned the basic usage of this reproducible workflow concept and there are a lot of ideas to improve my work based on that.
- working with workflow managers is essential in genomic studies and Nextflow display important functionalities that other workflow managers do not
- Because the team is well trained and able to address the questions.

Which topics were most useful or interesting to you and why?

- For me the presentation on how to use Nextflow was most useful
- DSL-2, As a new user to NextFlow and about to start my new project with BovReg in Jan 2021 knowing about DSL-2 and nf-core in advance is a key and I will start some detailed chat with the NF team closer to my start date.
- tutorial, because I did not know anything about nextflow
- SDL2 / to decide how to develop pipelines
- Nextflow examples & talks by Nextflow developers; directly applicable, while at the same time explaining why things are as they are.
- RNA-Seq analysis pipeline. Because it's in frame of my interests.
- Implementing the pipelines.
- All parts of workshop were useful.
- Nextflow training
- Practical training
- Nextflow best practices and nf-core layout
- The AWS and the RNAseq. because I am working at them
- overview of dsl2 was great to get a sense of how Nextflow is developing
- hands-on training and Nextflow Tower/nf-co.re because these were all very practical and useful information

- nf-core and next-flow Tower (seems to be a very useful solution for non-expert users to substitute for Galaxy) and the training to understand in detail the nextflow syntax and structure was also the most interesting for me.
- Everything was interesting, because I am a nextflow newbie. DSL2 was cool due to its big improvement step in structuring code. Extensive support is also important for novices via different channels (slack, github, youtube videos).
- Ewan and Paolo because it shows the basic understanding of the nextflow
- NF-core and Nextflow, more applicable to what I am working on

Which missing topics would you have liked to see covered?

- I think all beginners' relevant topics have been covered
- workflow documentation (writing help,...)
- Some more emphasis on practical details of running / configuring Nextflow with Docker
- Gene networking
- I am basically a computer science student, where I have no domain knowledge in biology or bioinformatics. I think I have learned a lot from basic stuff to implementing pipelines.
- How to use Dockers
- maybe more on integration with containers and the cloud
- Nothing to add ... first I need more experience in Nextflow.
- building up containers
- A bit more details of usage of containers in Nextflow process
- Containers, training thereof would be very useful