The Long-reads Transcriptomics European Consortium MSCA-DN Project (LongTREC): The make of

Clara Rodríguez, Project Manager
Institute for Integrative Systems Biology, CSIC, Spain
clara.rodriguez@csic.es

https://longtrec.eu/
@longtrec.bsky.social
https://www.linkedin.com/company/mscadnlongtrec

Ana Conesa, Research Professor
Institute for Integrative Systems Biology, CSIC, Spain
ana.conesa@csic.es

@anaconesa.bsky.social
@conesalab.bsky.social
https://www.linkedin.com/company/genomics-of-gene-expression-lab/

i2sysbio

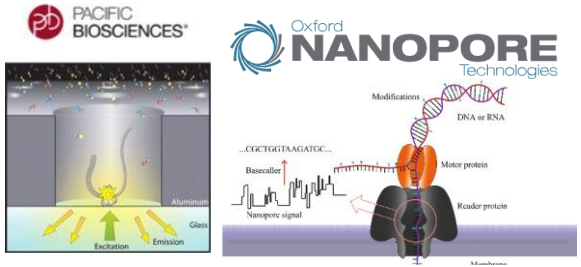
Genomics of Gene Expression Lab

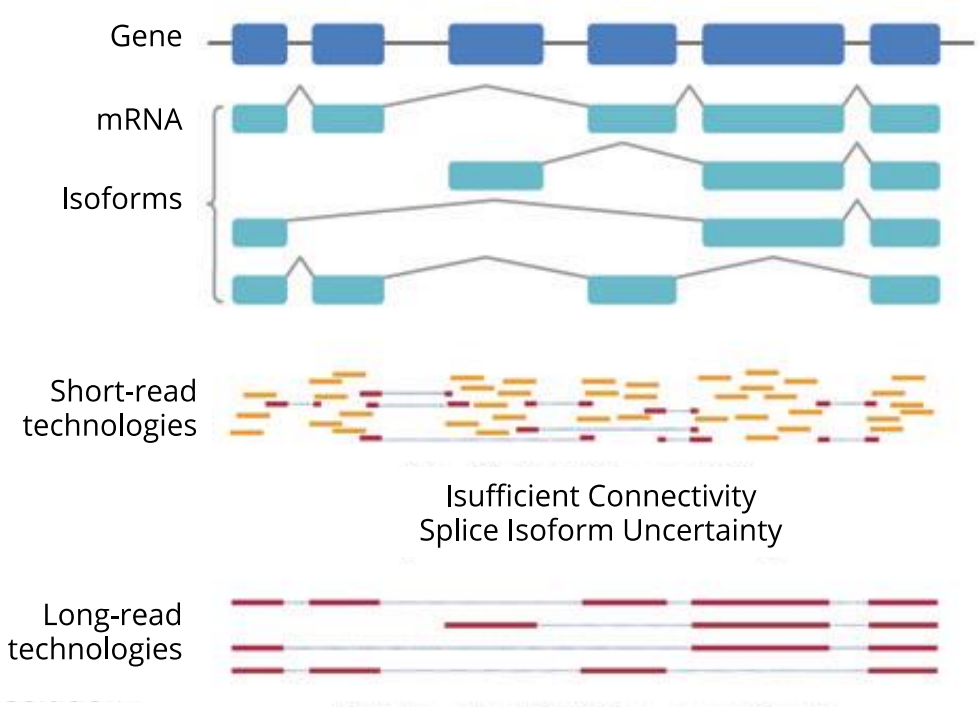
Successful key elements

- Clear, appealing and timely research question
- Participants are the EU leaders in the field
- Attractive reading and repetitive message
- Extremely complete training program

HIGH IMPACT (score 5/5)

The Science in highly timely





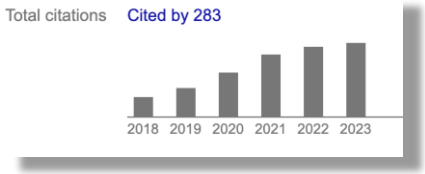
Ana Conesa's position in this science is high relevant

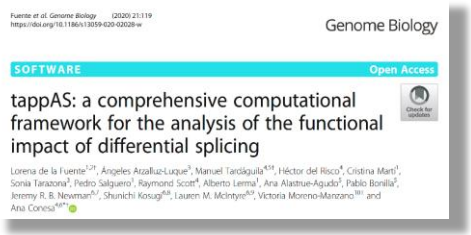
Method

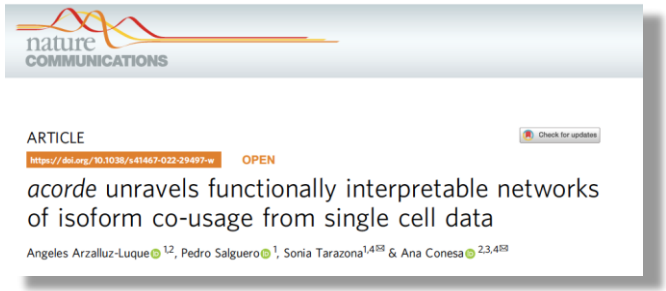
SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification

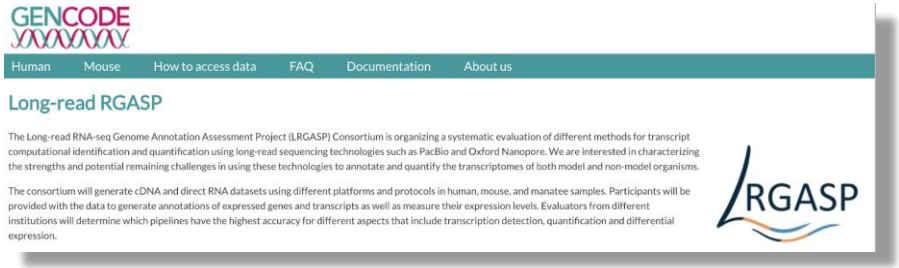
Manuel Tardaguila,[1,11] Lorena de la Fuente,[2,11] Cristina Marti,[2] Cécile Pereira,[1] Francisco Jose Pardo-Palacios,[2] Hector del Risco,[1] Marc Ferrell,[1] Maravillas Mellado,[3] Marissa Macchietto,[4] Kenneth Verheggen,[5,6] Mariola Edelmann,[1] Iakes Ezkurdia,[7] Jesus Vazquez,[7] Michael Tress,[8] Ali Mortazavi,[4] Lennart Martens,[5,6] Susana Rodriguez-Navarro,[9,10] Victoria Moreno-Manzano,[3] and Ana Conesa[1,2]

[1]Department of Microbiology and Cell Science, Institute for Food and Agricultural Sciences, Genetics Institute, University of Florida, Gainesville, Florida 32611, USA; [2]Genomics of Gene Expression Laboratory, Centro de Investigaciones Principe Felipe (CIPF), 46012 Valencia, Spain; [3]Neural Regeneration Laboratory, CIPF, 46012 Valencia, Spain; [4]Department of Developmental and Cell Biology, University of California, Irvine, California 92617, USA; [5]VIB-UGent Center for Medical Biotechnology, VIB, B-9000 Ghent, Belgium; [6]Department of Biochemistry, Ghent University, B-9000 Ghent, Belgium; [7]Centro Nacional de Investigaciones Cardiovasculares CNIC, 28029 Madrid, Spain; [8]Spanish National Cancer Research Centre (CNIO), 28029 Madrid, Spain; [9]Gene Expression and mRNA Metabolism Laboratory, CSIC, IBV, 46010 Valencia, Spain; [10]Gene Expression and mRNA Metabolism Laboratory, CIPF, 46012 Valencia, Spain









Successful key elements

- Clear, appealing and timely research question
- Participants are the EU leaders in the field
- Attractive reading and repetitive message
- Extremely complete training program

Supervisors: women and men, senior and junior academia and industry

Genomics of Gene Expression Lab

Additional Table 5. Supervisors (co indicates co-supervisor with)

Dr. Ana Conesa, CSIC-I2SysBio, co with Götz
Bioinformatician, Computational Biologist
137 publications; H-index 53; > 28,000 citations
Supervised 16 DCs, 15 ERs, 38 Masters

Dr. Kristina Gruden, NIB; co with Petek
Biologist and Biochemist
135 publications; H-index 45; > 5,600 citations
Supervised 25 DCs, 8 ERs, 12 Masters

Prof. Roderic Guigó, CRG, co with Novoa
Bioinformatician. **ERC awardee**
260 publications; H-index 115; > 125,000 citations
Supervised > 16 DCs, 25 ERs, 29 Masters

Dr. Pascal Barbry, CNRS
Biologist
30 Publications, H-index 67; > 19,000 citations
Supervised 28 DCs, 12 ERs, 12 Masters

Dr. Francesco Nicassio, IIT; co with Leonardi
RNA biologist, Cancer biologist
38 publications; H-index 21; > 2,100 citations
Supervised 10 DCs, 10 ERs, 5 Masters

Dr. Matt Loose, NU, co with Järvelin
Computation, Developmental biologist
48 Publications, H-index 37; > 8,000
Supervised 9 DCs, 5 Masters

Dr. Adam Ameur, UU, co with Sahlin
Bioinformatician, Biologist
71 publications; H-index 32; > 9,000 citations
Supervised 2 DCs, 4 Masters

Dr. Stefan Götz, BioBam
Bioinformatician, Computational Scientist
17 Publications, H-index: 13
Supervised 2 DCs; 5 Masters

Dr. Marko Petek, NIB; **Junior PI**
Computational Biologist
29 publications; H-index 15; > 700 citations
Supervised 1 DC (in progress),1 Masters

Dr. Eva María Novoa, CRG; **Junior PI**
Computational Biologist and Biochemist
30 Publications, H-index 13; > 1,200 citations
Supervised 2 DCs, 4 ERs, 2 Masters

Dr. Wilfried Haerty, EI
Bioinformatician, Computational Biologist
71 publications; H-index 29; > 7,000 citations
Supervised 3 DCs, 11 ERs, 0 Masters

Dr. Tommaso Leonardi, IIT; **Junior PI**
Computational Biologist, RNA biologist
24 publications; H-index 17; > 2,200 citations
Supervised 1 DC, 1ER 2, Master

Dr. Aino Järvelin, SU; **Junior PI**
Biochemist, Bioinformatician
20 Publications, H-index 14; > 1,100 citations
Supervised 4 DCs, 5 Masters

Dr. Kristoffer Sahlin, SU; **Junior PI**
Computer Scientist, Bioinformatician
14 Publications, H-index 9; > 1,500 citations
Supervised 1 DCs, 1 BSc, 1 Master's

Beneficiaries and Associated Partners

Genomics of Gene Expression Lab

- 9 beneficiaries
- 2 AAPP hiring DC (UK partners – own funding)
- 9 AAPP (secondments and/or training)
- 9 universities (AAPP) – only for the awarding PhD degree

 2 AAPP (secondments) left before the signature of the CA

Successful key elements

- Clear, appealing and timely research question
- Participants are the EU leaders in the field
- Attractive reading and repetitive message
- Extremely complete training program

LongTREC: research question

- The first page says it all.

1. Excellence

1.1 Quality and pertinence of the project's research and innovation objectives

1.1.1. Introduction, objectives, and overview of the research programme

Are we ready to embrace long reads technologies as the upcoming transcriptomics approach?

Over the last 10 years, a new third generation of sequencing (TGS) platforms, represented by Oxford Nanopore and Pacific Biosciences, have strongly emerged as novel sequencing methods. These technologies -in contrast to the traditional short-read Illumina- are able to sequence single molecules thanks to the utilization, respectively, of nanopores and microwells, and they can produce long (> tens of thousands of nucleotides) reads. These two properties - single molecule sequencing and long reads- create exciting opportunities for nucleic acid research. They not only resolve long-range phasing problems intractable to the short reads, but they also couple sequencing to the detection of nucleotide modifications such as methylation, resulting into a simultaneous measurement of the genome and the epigenome. Applied to transcriptome research, long-read RNA sequencing (lrRNA-seq) has important advantages. Long read sequencing (LRS) can obtain full-length transcripts, making possible the accurate detection of isoforms, alleles, and transcripts from gene families, traditionally hard to resolve by short reads. Additionally, the Nanopore technology can perform direct sequencing of RNA molecules (dRNA-seq) and capture their nucleotide modifications, enabling new venues for epitranscriptomics research (**Figure 1**). LRS have already made breakthroughs such as the completion of the human genome from telomere to telomere[1] and is the driving technology of the Vertebrate Genome and Earth BioGenome projects[2]. Today, LRS has replaced short reads in genome assembly, and it is generally accepted by the community that no genome sequencing project is further conceived without the utilization of LRS. **We anticipate that, in the same way that LRS has conquered the sequencing of genomes, lrRNA-seq will be the preferred transcriptome profiling technology in the near future.**

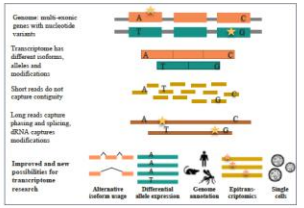


Figure 1. The power of LRS for transcriptome analysis.

However, lrRNA-sequencing is not an error-free technology. Limitations still exist at the library preparation, mapping, transcript reconstruction and quantifications steps. Upcoming applications such as single-cell sequencing, or allele-specific expression require well established protocols and benchmarked analysis pipelines. Additionally, the power of LRS to dissect the regulatory biology of RNAs needs to be further explored. In the LongTREC MSCA-DN we propose to address these challenges by collaborative research and training of **10 Doctoral Candidates** (DC) who will acquire the set of competences to achieve this transformation. Our major Research Objectives are:

Objective 1:	Improve library preparation, sequencing, and pre-processing steps to obtain better lrRNA-seq data.
Objective 2:	Develop lrRNA-seq methods and pipelines to address pending questions in transcriptome biology.
Objective 3:	Advance the LRS transcriptomics field with multi-omics applications.
Objective 4:	Create community-wide and user-friendly tools for strong impact.

To make this ambitious goal possible we have gathered the current European leaders of the lrRNA-seq technologies, both at experimentation and computation (see beneficiary info). **Ana Conesa** (CISC-I2Sysbio) was pioneer in the development of SQANTI, a lrRNA-seq quality control (QC) tool, now considered a reference software among the lrRNA-seq community. Furthermore, she is co-organizer of the LRGASP initiative, a community-wide challenge to evaluate the potential of LRS for transcriptome analyses[3]. **Kristoffer Sahlin** (SU) brings expertise in LR mapping and sequence analysis. **Wilfried Hearty** (EI) is a cell biologist who uses LRS to understand cell-state and the impact of mutations on splicing. **Roderic Guigó** (CRG) is a world-expert in genome annotation, member of GENCODE and developer of the CapTrap method for 5'capture of RNAs. **Pascal Barbry** (CNRS) is an expert in genomic technologies who has develop tools for single-cell analysis using long reads. **Kristina Gruden** and **Marko Petek** (NIB) published the potato pan-transcriptome annotation using LRS and bring expertise on polyploid genomes. Both **Tommaso Leonardi** (IIT) and **Eva Maria Novoa** (CRG) are rising stars in epitranscriptomics, while **Francesco Nicassio** is an expert in microRNAs. Finally, we count with **Nanopore**, the European Leader in LRS who will deliver improved sequencing and with **Biobam**, a bioinformatics company who will bring the LRS methods to the biologists.

LongTREC: research question

- The first page says it all.
- Actually, the first sentence says it all!

1. Excellence

1.1.1. Introduction, objectives, and overview of the research programme

Are we ready to embrace long reads technologies as the upcoming transcriptomics approach?

Pacific Biosciences, have strongly emerged as novel sequencing methods. These technologies -in contrast to the traditional short-read Illumina- are able to sequence single molecules thanks to the utilization, respectively, of nanopores and microwells, and they can produce long (> tens of thousands of nucleotides) reads. These two properties - single molecule sequencing and long reads- create exciting opportunities for nucleic acid research. They not only resolve long-range phasing problems intractable to the short reads, but they also couple sequencing to the detection of nucleotide modifications such as methylation, resulting into a simultaneous measurement of the genome and the epigenome. Applied to transcriptome research, long-read RNA sequencing (lrRNA-seq) has important advantages. Long read sequencing (LRS) can obtain full-length transcripts, making possible the accurate detection of isoforms, alleles, and transcripts from gene families, traditionally hard to resolve by short reads. Additionally, the Nanopore technology can perform direct sequencing of RNA molecules (dRNA-seq) and capture their nucleotide modifications, enabling new venues for epitranscriptomics research (**Figure 1**). LRS have already made breakthroughs such as the completion of the human genome from telomere to telomere[1] and is the driving technology of the Vertebrate Genome and Earth BioGenome projects[2]. Today, LRS has replaced short reads in genome assembly, and it is generally accepted by the community that no genome sequencing project is further conceived without the utilization of LRS. **We anticipate that, in the same way that LRS has conquered the sequencing of genomes, lrRNA-seq will be the preferred transcriptome profiling technology in the near future.**

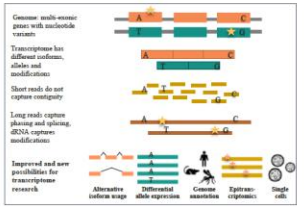


Figure 1. The power of LRS for transcriptome analysis.

However, lrRNA-sequencing is not an error-free technology. Limitations still exist at the library preparation, mapping, transcript reconstruction and quantifications steps. Upcoming applications such as single-cell sequencing, or allele-specific expression require well established protocols and benchmarked analysis pipelines. Additionally, the power of LRS to dissect the regulatory biology of RNAs needs to be further explored. In the LongTREC MSCA-DN we propose to address these challenges by collaborative research and training of **10 Doctoral Candidates** (DC) who will acquire the set of competences to achieve this transformation. Our major Research Objectives are:

Objective 1:	Improve library preparation, sequencing, and pre-processing steps to obtain better lrRNA-seq data.
Objective 2:	Develop lrRNA-seq methods and pipelines to address pending questions in transcriptome biology.
Objective 3:	Advance the LRS transcriptomics field with multi-omics applications.
Objective 4:	Create community-wide and user-friendly tools for strong impact.

To make this ambitious goal possible we have gathered the current European leaders of the lrRNA-seq technologies, both at experimentation and computation (see beneficiary info). **Ana Conesa** (CISC-I2Sysbio) was pioneer in the development of SQANTI, a lrRNA-seq quality control (QC) tool, now considered a reference software among the lrRNA-seq community. Furthermore, she is co-organizer of the LRGASP initiative, a community-wide challenge to evaluate the potential of LRS for transcriptome analyses[3]. **Kristoffer Sahlin** (SU) brings expertise in LR mapping and sequence analysis. **Wilfried Hearty** (EI) is a cell biologist who uses LRS to understand cell-state and the impact of mutations on splicing. **Roderic Guigó** (CRG) is a world-expert in genome annotation, member of GENCODE and developer of the CapTrap method for 5'capture of RNAs. **Pascal Barbry** (CNRS) is an expert in genomic technologies who has develop tools for single-cell analysis using long reads. **Kristina Gruden** and **Marko Petek** (NIB) published the potato pan-transcriptome annotation using LRS and bring expertise on polyploid genomes. Both **Tommaso Leonardi** (IIT) and **Eva Maria Novoa** (CRG) are rising stars in epitranscriptomics, while **Francesco Nicassio** is an expert in microRNAs. Finally, we count with **Nanopore**, the European Leader in LRS who will deliver improved sequencing and with **Biobam**, a bioinformatics company who will bring the LRS methods to the biologists.

LongTREC: research question

- The first page says it all.
- Actually, the first sentence says it all!
- The first figure illustrates the concept

1. Excellence

1.1 Quality and pertinence of the project's research and innovation objectives

1.1.1. Introduction, objectives, and overview of the research programme

Are we ready to embrace long re

Over the last 10 years, a new third
Pacific Biosciences, have strongly
traditional short-read Illumina- ar
nanopores and microwells, and the
of thousands of nucleotides) read
single molecule sequencing and l
opportunities for nucleic acid resea
long-range phasing problems intr
but they also couple sequencing to
modifications such as methyl
simultaneous measurement of the g
Applied to transcriptome res
sequencing (lrRNA-seq) has impor
sequencing (LRS) can obtain full-
possible the accurate detection
transcripts from gene families, trad
short reads. Additionally, the N
perform direct sequencing of RNA
capture their nucleotide modificati
already made breakthroughs such
driving technology of the Vertebra
in genome assembly, and it is gen
conceived without the utilization
sequencing of genomes, lrRNA-s

However, lrRNA-sequencing is r
mapping, transcript reconstruction
or allele-specific expression requi
the power of LRS to dissect the re

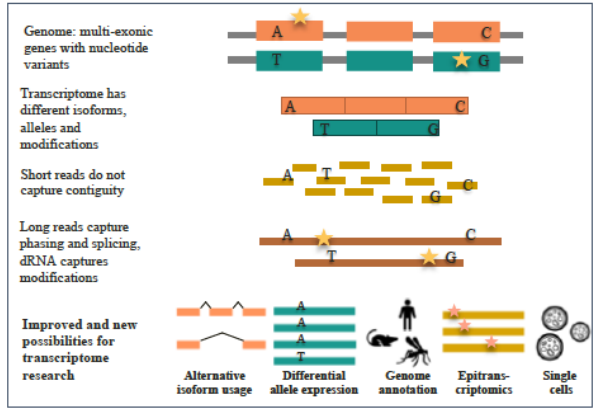


Figure 1. The power of LRS for transcriptome analysis.

DN we propose to address these challenges by collaborative research and training of **10 Doctoral Candidates** (DC) who will acquire the set of competences to achieve this transformation. Our major Research Objectives are:

Objective 1:	Improve library preparation, sequencing, and pre-processing steps to obtain better lrRNA-seq data.
Objective 2:	Develop lrRNA-seq methods and pipelines to address pending questions in transcriptome biology.
Objective 3:	Advance the LRS transcriptomics field with multi-omics applications.
Objective 4:	Create community-wide and user-friendly tools for strong impact.

To make this ambitious goal possible we have gathered the current European leaders of the lrRNA-seq technologies, both at experimentation and computation (see beneficiary info). **Ana Conesa** (CISC-I2Sysbio) was pioneer in the development of SQANTI, a lrRNA-seq quality control (QC) tool, now considered a reference software among the lrRNA-seq community. Furthermore, she is co-organizer of the LRGASP initiative, a community-wide challenge to evaluate the potential of LRS for transcriptome analyses[3]. **Kristoffer Sahlin** (SU) brings expertise in LR mapping and sequence analysis. **Wilfried Hearty** (EI) is a cell biologist who uses LRS to understand cell-state and the impact of mutations on splicing. **Roderic Guigó** (CRG) is a world-expert in genome annotation, member of GENCODE and developer of the CapTrap method for 5'capture of RNAs. **Pascal Barbry** (CNRS) is an expert in genomic technologies who has develop tools for single-cell analysis using long reads. **Kristina Gruden** and **Marko Petek** (NIB) published the potato pan-transcriptome annotation using LRS and bring expertise on polyploid genomes. Both **Tommaso Leonardi** (IIT) and **Eva Maria Novoa** (CRG) are rising stars in epitranscriptomics, while **Francesco Nicassio** is an expert in microRNAs. Finally, we count with **Nanopore**, the European Leader in LRS who will deliver improved sequencing and with **Biobam**, a bioinformatics company who will bring the LRS methods to the biologists.

LongTREC: research question

- The first page says it all.
- Actually, the first sentence says it all!
- The first figure illustrates the concept
- Refer to an accepted hypothesis to create a timely question

1. Excellence

1.1 Quality and pertinence of the project's research and innovation objectives

1.1.1. Introduction, objectives, and overview of the research programme

Are we ready to embrace long reads technologies as the upcoming transcriptomics approach?

Over the last 10 years, a new third generation of sequencing (TGS) platforms, represented by Oxford Nanopore and Pacific Biosciences, have strongly emerged as novel sequencing methods. These technologies -in contrast to the traditional short-read Illumina- are able to sequence single molecules thanks to the utilization, respectively, of nanopores and microwells, and they can produce long (> tens of thousands of nucleotides) reads. These two properties - single molecule sequencing and long reads- create exciting opportunities for nucleic acid research. They not only resolve long-range phasing problems intractable to the short reads, but they also couple sequencing to the detection of nucleotide modifications such as methylation, resulting into a simultaneous measurement of the genome and the epigenome. Applied to transcriptome research, long-read RNA sequencing (lrRNA-seq) has important advantages. Long read sequencing (LRS) can obtain full-length transcripts, making possible the accurate detection of isoforms, alleles, and transcripts from gene families, traditionally hard to resolve by short reads. Additionally, the Nanopore technology can perform direct sequencing of RNA molecules (dRNA-seq) and capture their nucleotide modifications, enabling new venues for epitranscriptomics research (**Figure 1**). LRS have already made breakthroughs such as the completion of the human genome from telomere to telomere[1] and is the driving technology of the Vertebrate Genome and Earth BioGenome projects[2]. Today, LRS has replaced short reads in genome assembly, and it is generally accepted by the community that no genome sequencing project is further conceived without the utilization of LRS. **We anticipate that, in the same way that LRS has conquered the sequencing of genomes, lrRNA-seq will be the preferred transcriptome profiling technology in the near future.**

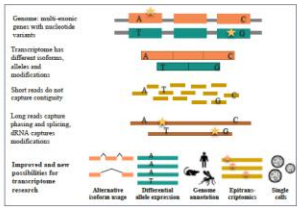


Figure 1. The power of LRS for transcriptome analysis.

However, lrRNA sequencing is not an error-free technology. Limitations still exist at the library preparation

To make this ambitious goal possible we have gathered the current European leaders of the lrRNA-seq technologies, both at experimentation and computation (see beneficiary info). **Ana Conesa** (CISC-I2Sysbio) was pioneer in the development of SQANTI, a lrRNA-seq quality control (QC) tool, now considered a reference software among the lrRNA-seq community. Furthermore, she is co-organizer of the LRGASP initiative, a community-wide challenge to evaluate the potential of LRS for transcriptome analyses[3]. **Kristoffer Sahlin** (SU) brings expertise in LR mapping and sequence analysis. **Wilfried Hearty** (EI) is a cell biologist who uses LRS to understand cell-state and the impact of mutations on splicing. **Roderic Guigó** (CRG) is a world-expert in genome annotation, member of GENCODE and developer of the CapTrap method for 5'capture of RNAs. **Pascal Barbry** (CNRS) is an expert in genomic technologies who has develop tools for single-cell analysis using long reads. **Kristina Gruden** and **Marko Petek** (NIB) published the potato pan-transcriptome annotation using LRS and bring expertise on polyploid genomes. Both **Tommaso Leonardi** (IIT) and **Eva Maria Novoa** (CRG) are rising stars in epitranscriptomics, while **Francesco Nicassio** is an expert in microRNAs. Finally, we count with **Nanopore**, the European Leader in LRS who will deliver improved sequencing and with **Biobam**, a bioinformatics company who will bring the LRS methods to the biologists.

LongTREC: research question

- The first page says it all.
- Actually, the first sentence says it all!
- The first figure illustrates the concept
- Refer to an accepted hypothesis to create a timely question
- Clear objectives in first page

1. Excellence

1.1 Quality and pertinence of the project's research and innovation objectives

1.1.1. Introduction, objectives, and overview of the research programme

Are we ready to embrace long reads technologies as the upcoming transcriptomics approach?

Over the last 10 years, a new third generation of sequencing (TGS) platforms, represented by Oxford Nanopore and Pacific Biosciences, have strongly emerged as novel sequencing methods. These technologies -in contrast to the traditional short-read Illumina- are able to sequence single molecules thanks to the utilization, respectively, of nanopores and microwells, and they can produce long (> tens of thousands of nucleotides) reads. These two properties - single molecule sequencing and long reads- create exciting opportunities for nucleic acid research. They not only resolve long-range phasing problems intractable to the short reads, but they also couple sequencing to the detection of nucleotide modifications such as methylation, resulting into a simultaneous measurement of the genome and the epigenome. Applied to transcriptome research, long-read RNA sequencing (lrRNA-seq) has important advantages. Long read sequencing (LRS) can obtain full-length transcripts, making possible the accurate detection of isoforms, alleles, and transcripts from gene families, traditionally hard to resolve by short reads. Additionally, the Nanopore technology can perform direct sequencing of RNA molecules (dRNA-seq) and capture their nucleotide modifications, enabling new venues for epitranscriptomics research (**Figure 1**). LRS have already made breakthroughs such as the completion of the human genome from telomere to telomere[1] and is the driving technology of the Vertebrate Genome and Earth BioGenome projects[2]. Today, LRS has replaced short reads in genome assembly, and it is generally accepted by the community that no genome sequencing project is further conceived without the utilization of LRS. **We anticipate that, in the same way that LRS has conquered the sequencing of genomes, lrRNA-seq will be the preferred transcriptome profiling technology in the near future.**

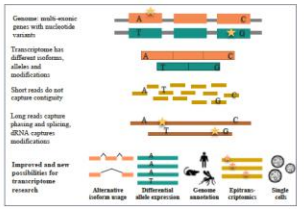


Figure 1. The power of LRS for transcriptome analysis.

However, lrRNA-sequencing is not an error-free technology. Limitations still exist at the library preparation, mapping, transcript reconstruction and quantifications steps. Upcoming applications such as single-cell sequencing, or allele-specific expression require well established protocols and benchmarked analysis pipelines. Additionally, the power of LRS to dissect the regulatory biology of RNAs needs to be further explored. In the LongTREC MSCA

Objective 1: Improve library preparation, sequencing, and pre-processing steps to obtain better lrRNA-seq data.
Objective 2: Develop lrRNA-seq methods and pipelines to address pending questions in transcriptome biology.
Objective 3: Advance the LRS transcriptomics field with multi-omics applications.
Objective 4: Create community-wide and user-friendly tools for strong impact.

development of SQANTI, a lrRNA-seq quality control (QC) tool, now considered a reference software among the lrRNA-seq community. Furthermore, she is co-organizer of the LRGASP initiative, a community-wide challenge to evaluate the potential of LRS for transcriptome analyses[3]. **Kristoffer Sahlin** (SU) brings expertise in LR mapping and sequence analysis. **Wilfried Hearty** (EI) is a cell biologist who uses LRS to understand cell-state and the impact of mutations on splicing. **Roderic Guigó** (CRG) is a world-expert in genome annotation, member of GENCODE and developer of the CapTrap method for 5'capture of RNAs. **Pascal Barbry** (CNRS) is an expert in genomic technologies who has develop tools for single-cell analysis using long reads. **Kristina Gruden** and **Marko Petek** (NIB) published the potato pan-transcriptome annotation using LRS and bring expertise on polyploid genomes. Both **Tommaso Leonardi** (IIT) and **Eva Maria Novoa** (CRG) are rising stars in epitranscriptomics, while **Francesco Nicassio** is an expert in microRNAs. Finally, we count with **Nanopore**, the European Leader in LRS who will deliver improved sequencing and with **Biobam**, a bioinformatics company who will bring the LRS methods to the biologists.

LongTREC: research question

- The first page says it all.
- Actually, the first sentence says it all!
- The first figure illustrates the concept
- Refer to an accepted hypothesis to create a timely question
- Clear objectives in first page
- And a dream team, all with high impact publications in Long Reads

1. Excellence

1.1 Quality and pertinence of the project's research and innovation objectives

1.1.1. Introduction, objectives, and overview of the research programme

Are we ready to embrace long reads technologies as the upcoming transcriptomics approach?

Over the last 10 years, a new third generation of sequencing (TGS) platforms, represented by Oxford Nanopore and Pacific Biosciences, have strongly emerged as novel sequencing methods. These technologies -in contrast to the traditional short-read Illumina- are able to sequence single molecules thanks to the utilization, respectively, of nanopores and microwells, and they can produce long (> tens of thousands of nucleotides) reads. These two properties - single molecule sequencing and long reads- create exciting opportunities for nucleic acid research. They not only resolve long-range phasing problems intractable to the short reads, but they also couple sequencing to the detection of nucleotide modifications such as methylation, resulting into a simultaneous measurement of the genome and the epigenome. Applied to transcriptome research, long-read RNA sequencing (lrRNA-seq) has important advantages. Long read sequencing (LRS) can obtain full-length transcripts, making possible the accurate detection of isoforms, alleles, and transcripts from gene families, traditionally hard to resolve by short reads. Additionally, the Nanopore technology can perform direct sequencing of RNA molecules (dRNA-seq) and capture their nucleotide modifications, enabling new venues for epitranscriptomics research (**Figure 1**). LRS have already made breakthroughs such as the completion of the human genome from telomere to telomere[1] and is the driving technology of the Vertebrate Genome and Earth BioGenome projects[2]. Today, LRS has replaced short reads in genome assembly, and it is generally accepted by the community that no genome sequencing project is further conceived without the utilization of LRS. **We anticipate that, in the same way that LRS has conquered the sequencing of genomes, lrRNA-seq will be the preferred transcriptome profiling technology in the near future.**

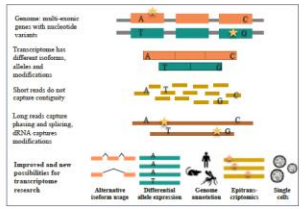


Figure 1. The power of LRS for transcriptome analysis.

However, lrRNA-sequencing is not an error-free technology. Limitations still exist at the library preparation, mapping, transcript reconstruction and quantifications steps. Upcoming applications such as single-cell sequencing, or allele-specific expression require well established protocols and benchmarked analysis pipelines. Additionally, the power of LRS to dissect the regulatory biology of RNAs needs to be further explored. In the LongTREC MSCA-

To make this ambitious goal possible we have gathered the current European leaders of the lrRNA-seq technologies, both at experimentation and computation (see beneficiary info). **Ana Conesa** (CISC-I2Sysbio) was pioneer in the development of SQANTI, a lrRNA-seq quality control (QC) tool, now considered a reference software among the lrRNA-seq community. Furthermore, she is co-organizer of the LRGASP initiative, a community-wide challenge to evaluate the potential of LRS for transcriptome analyses[3]. **Kristoffer Sahlin** (SU) brings expertise in LR mapping and sequence analysis. **Wilfried Hearty** (EI) is a cell biologist who uses LRS to understand cell-state and the impact of mutations on splicing. **Roderic Guigó** (CRG) is a world-expert in genome annotation, member of GENCODE and developer of the CapTrap method for 5'capture of RNAs. **Pascal Barbry** (CNRS) is an expert in genomic technologies who has develop tools for single-cell analysis using long reads. **Kristina Gruden** and **Marko Petek** (NIB) published the potato pan-transcriptome annotation using LRS and bring expertise on polyploid genomes. Both **Tommaso Leonardi** (IIT) and **Eva Maria Novoa** (CRG) are rising stars in epitranscriptomics, while **Francesco Nicassio** is an expert in microRNAs. Finally, we count with **Nanopore**, the European Leader in LRS who will deliver improved sequencing and with **Biobam**, a bioinformatics company who will bring the LRS methods to the biologists.

The message is constantly and consistently repeated

- *We are going to develop the software to make lrRNA-seq the transcriptomics technology of the near future* (High impact project)

All the above reveals that lrRNA-seq is an effervescent field, where new experimental protocols, biological applications and analysis solutions are constantly emerging in pace with the rapid improvements in third-generation sequencing technologies. LrRNA-seq has already displayed great potential, but much work is still needed. A new generation of computational biologists is essential to consolidate current methods, develop new tools, envision new applications, and lead the unstoppable transition of transcriptome analysis to effectively become LRS-based. **The ambition of LongTREC is that this new generation of scientific leaders is European**.

is also related to WP1. Similarly, DC8

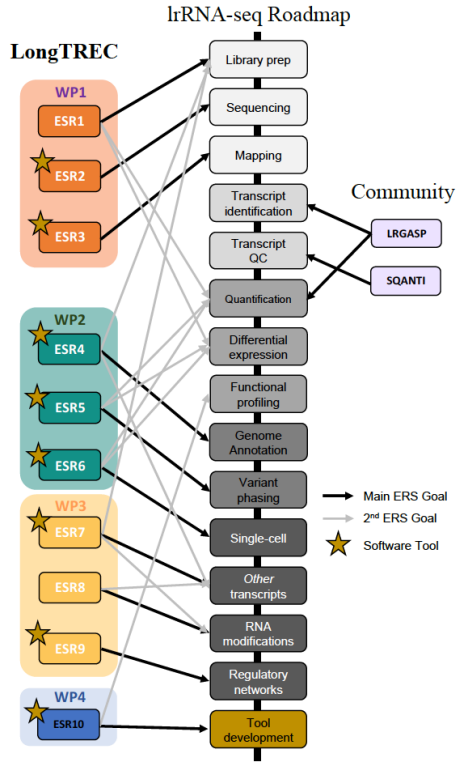


Figure 2. LongTREC in the lrRNA-seq Roadmap

Present the must-read items as tables

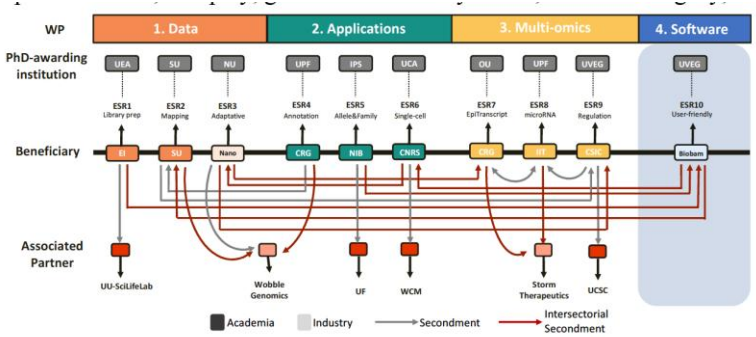


Figure 4. LongTREC secondments plan.

Target employer	Specific skills trained by LongTREC
Research Institution	Comprehensive SSV training. Training in grant (both EU and USA styles) writing. Global academic networking.
Academia, teachers	Preparation of educational materials. Teaching at LongTREC summer school.
Spin-offs / Start-ups	Entrepreneurship, research skills, project management. Role models by LongTREC partners.
Industry	Project management, leadership, communication and networking capacity. Training in corporate values.
Health sector	State-of-the-art knowledge of the application of genomics technologies in cancer and developmental diseases. Bioinformatics and biostatistics for clinical research. Training in General Data Protection Regulations.
Editorials	Scientific knowledge, writing and reviewing skills. Organisation of a Special Collection of papers.
Consulting	Project management, feasibility assessment, communication skills, entrepreneurship. Code of conduct training.
Funding agencies	EU funding, GDPR, animal regulations. Ethics training. Project management. Peer-review of project proposals.
Science & Politics	Communication skills, Ethics, Science Diplomacy (lecture by Dr Capua: virologist, and politician).
NGOs	RRI, communication, project management, research knowledge. Lecture from NGO funding research.
Global organisations	SSV training, transversal knowledge. Lectures & interaction with leaders of ISCB, Goblet, ELIXIR, EMBO, EBI.

Additional Table 9. LongTREC training improved employability for specific employers.

2.1.2 Contribution by the private sector. They contribute to LongTREC DCs training at different levels:

Beneficiaries (B) Partners (P)		Expertise	DC Home or Secondm	Network-wide training* ❸-❷ Science ❶-❸ Skills
NANOPORE	B	World-leader in Third Generation Sequencing Technologies	DC2 (H) DC6 (S) DC9 (S)	❶ Technology Lecture ❸ Translation Lecture. ● Leader of library preparation research (WP1) ● Contribute to educational materials
biobam	B	Creation of user-friendly bioinformatics software	DC10 (H) DC1 (S) DC8 (S)	❸ Translation Lecture ❶ Careers Lecture ❷ Exploitation Lecture ● Significant contribution to programming training ● Contribute to educational materials
Wobble GENOMICS	P	Innovation in the use of DNA and RNA as biomarkers	DC5 (S) DC6 (S)	❶ Technology Lecture ❸ Translation Lecture ❷ Exploitation Lecture
STORM THERAPEUTICS	P	Targeting RNA modifying enzymes to tackling cancer	DC7 (S), DC8 (S)	❸ Translation Lecture ❶ Careers Lecture
sb science management	P	Consultants for innovation management, standardization / data management	n.a.	❹ Will deliver the entire management training in two blocks of three days. Provide tools for conflict management

Additional Table 8. Training by the non-academic sector. *Please see also Tables 1.3b
*Moreover, professionals from Roche (Pharmaceutical), ADM (Food and Nutrition), IVI (In-vitro fertilization), BASF (Chemicals), Igenomix (Reproductive Genomics), Sequentia Biopharma (Bioinformatics) and INMEGEN (Biotechnology) will contribute with a diversity of specialized Lectures in ❷ Translation ❶ Careers and ❹ Values.

Additional Table 7. Contributions

		Beneficiary									Associated Partners									
	WP	CSIC	SU	CNRS	IIT	NIB	CRG	EI	Nanopore	Biobam	BSC-Elixir	GTPB	GOBLET	UU	UCSC	UF	WCM	Wobble	SBMS	Strom	UN
Work package leader		◆		◆		◆	◆		◆	◆											
Host ESR	**All**	◆	◆	◆	◆	◆	◆	◆	◆	◆											
Secondment	**All**	◆	◆	◆	◆		◆		◆	◆				◆	◆	◆	◆	◆		◆	
Mentoring Team	**All**	◆	◆	◆	◆	◆	◆	◆	◆	◆				◆							◆
LRS Technology	**1-4**	◆	◆	◆	◆	◆	◆	◆	◆	◆				◆	◆	◆	◆				
Data Science	**1-4**	◆		◆											◆	◆					
Software Science	**1-4**	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆										◆
Local Training	**5**	◆	◆	◆	◆	◆	◆	◆	◆	◆											◆
Translation	**5**	◆	◆	◆	◆	◆	◆	◆	◆	◆								◆		◆	
Career Orientation	**5**	◆	◆	◆	◆	◆	◆	◆	◆	◆								◆		◆	◆
Transversal Skills	**5**	◆	◆	◆	◆		◆	◆	◆	◆		◆							◆		
Dissemination	**6**	◆	◆	◆	◆	◆	◆	◆	◆	◆		◆	◆							◆	

WP contains the contributing doctoral projects

Genomics of Gene Expression Lab

Table 3.1a Description of Work Packages

WP 1	**Improved methods for lrRNA-seq data acquisition and preparation** lead by Jarveling (Nanopore)	**Months 7 – 48**
Objectives Develop new experimental and bioinformatics methods that improve LRS data acquisition and pre-processing to deliver data structures that can successfully be utilized in downstream analyses. We work at the 3 mains steps of the pre-processing workflow: Library preparation (***Obj 1.1***), sequencing (***Obj 1.2***) and mapping of reads to a reference genome (***Obj 1.3***).		
Description of Work and Roles. Different library preparation efforts will target optimized RNA capture for different sample scenarios. ***Task 1.1***. Pacbio library preparation for low input material using a transcript concatenation strategy (DC1 + SciLifeLab learn multiple Pacbio library preparation methods) using hematopoietic cells. ***Task 1.2***. Pacbio library preparation to capture very long transcripts (DC4 + Wobble) in mammalian transcriptomes. ***Task 1.3***. Nanopore direct RNA library preparation for non-polyadenylated transcripts based on template switching approach (DC7 + Nanopore). Sperm RNA will be used. ***Task 1.4***. Nanopore direct RNA library preparation to capture pri-microRNAs (DC8). ENCODE cell lines will be used. ***Task 1.5***. To develop an adaptative sampling sequencing for Nanopore dRNA-seq and use to characterize lncRNAs (Nanopore). Tasks 1-5 results will be compared to current approaches and evaluated with SQANTI. ***Task 1.6*** To develop improved mapping algorithms based on strobemers that overcome the limitations of the uLTRA mapper to deal with novel transcripts (*uLTRAv2)* and obtain transcript models in the absence of a genome reference (*uLTRAv3)* (SU). This task will utilize real and simulated data and compare to state-of-the-art methods.		
Deliverables. *D1.1.* Progress report on the state of library preparation, sequencing, and mapping algorithm developments (Month 18). ***D1.2.*** Updated report on the state of library preparation, sequencing and mapping algorithm developments, including links to a protocol repository (i.e. protocols.io) and GitHub for code of novel mapping algorithms (Month 30). ***D1.3*** Report with an updated list of links to a repository containing novel lrRNA-seq library prep. protocols for and links to GitHub for uLTRAv2 and uLTRAv3 (Month 48).		

Each doctoral project is actually very short

Genomics of Gene Expression Lab

DC9	CSIC	University of Valencia	6	36	D3.1;D3.2;D3.3
Long reads multiomics methods to understand isoform regulatory biology (WP3)					
Objectives: Transcript variant expression is the combination of transcriptional and post-transcriptional regulation. The regulatory environment at promotor regions is studied by short reads, but these do not reveal the contiguous promoter methylation landscape, which could be captured by LRS. Moreover, methods for integration of multi-omics long reads data do not exist. Our objectives are: ◉ Develop a wet-lab protocol for long reads ATAC-seq and Nascent-seq using Nanopore. Apply to H1 human cell line. ◉ Develop an analysis pipeline to process lrATAC-seq data to identify chromatin accessibility and methylation. ◉ Integrate H1 lr-ATAC-seq and Nascent-seq data with LRGASP lrRNA-seq to infer transcript-specific regulatory patterns. ◉ Implement methods in the SQANTI and tappAS software tools.					
Expected Results: ® Long reads multi-omics dataset for H1 cell-line. ® Analysis pipeline for the integration of long reads multi-omics data. ® Model of isoform-specific transcriptional regulation. **Contingency:** © lrATAC-seq or lrNascent-seq library preps fail, contact developing lab (@ UF, that is an AP) or use simulated data. © No significant isoform regulatory signals from LRS data, evaluate short-read ATAC-seq from same ENCODE lines and confirm or reject failure.					
Planned secondment(s): ITT; Nicassio/Leornardi: Develop lrATAC-seq assay (Month 8, 2 months). Nanopore; Aino Javelin: Develop lrATAC-seq assay (Month 12, 2 months). UCSC; Angela Brooks: Learn FLAIR to pre-process Nanopore data (Month 24, 3 months).					

Deliverables and risks deal with forms and management

Genomics of Gene Expression Lab

Table 3.1b Deliverables List (ordered by WP)

#	Deliverable Title	WP	Lead Benefic.	Type	Level	Due
Scientific Deliverables						
D1.1	Progress report on the state of library preparation, sequencing and mapping algorithm	1	EI	R	PU	18
D1.2	Interim report including links to a protocol repository	1	EI	R	PU	30
D1.3	Final report with complete list of links to library preparation protocols	1	EI	R	PU	48
D2.1	Initial progress report on the state of algorithm developments in WP2	2	NIB	R	PU	18
D2.2	Interim report on the state of algorithm developed in WP2, including benchmarks	2	NIB	R	PU	30
D2.3	Final report with description of validated algorithms developed in WP2	2	NIB	R	PU	48
D3.1	Progress report on the state of algorithm developments in WP3	3	CRG	R	PU	18
D3.2	Interim report on the state of algorithm developments in WP3 applied to use cases	3	CRG	R	PU	30
D3.3	Final report with description of validated algorithms developed in WP3 & biological results	3	CRG	R	PU	48
D4.1	Progress report on the state of algorithm developments in WP4	4	CSIC	R	PU	18
D4.2	Interim report on the state of algorithm developed in WP4 and pipelines comparison	4	CSIC	R	PU	30
D4.3	Final report with complete list of LongTREC software and links to open repositories	4	CSIC	R	PU	48
Management, Training, Recruitment and Dissemination Deliverables						
D5.1	Initial Training Report including DC enrolment and TABs composition	5	CNRS	R	CO	18
D5.2	Document collecting the Career Development Plans	5	CNRS	R	CO	24
D5.3	Updated Training Report including Career Development Plans	5	CNRS	R	CO	30
D5.4	Updated Training Report including materials for Summer School and PhD defense status	5	CNRS	R	CO	48
D6.1	LongTREC website and social network profiles	7	EI	R	PU	2
D6.2	LongTREC dissemination and communication plan	7	EI	R	CO	10
D6.3	Report about the LongTREC Summer School	7	EI	R	PU	36
D6.4	Report on of all dissemination, public engagement, and communication activities	7	EI	R	CO	48
D7.1	Briefing notes for kick-off meeting delivered to all Supervisors	8	CSIC	R	CO	1
D7.2	Agreed operational procedures	8	CSIC	R	CO	2
D7.3	Completed Declarations on the Conformity for recruitments of DC position	8	CSIC	R	CO	10
D7.4	Data Management Plan (DMP) considering all LongTREC datasets	8	CSIC	R	CO	12
D7.5	Year 1 progress report to the REA	8	CSIC	R	CO	12
D7.6	Mid-term meeting, organized with the grant authority	8	CSIC	R	CO	24
D7.7	Year 3 reports to the REA	8	CSIC	R	CO	36
D7.8	Year 4 reports to the REA, including updates in the DMP and dissemination plan	8	CSIC	R	CO	48
D7.9	Evaluation questionary completed by each recruited DC	8	CSIC	R	CO	48

Table 3.1e Implementation Risks

	Description of Risk	Proposed mitigation measure
R1	The selected candidate renounces the offer	The next candidate from the list of interviewees will be selected.
R2	The fellow wants to change their PhD subject	The Supervisor and mentoring team will discuss it and take the matter to the Supervisory Board, if necessary.
R3	The fellow leaves during the fellowship	The contract is stopped and the following in the list of candidates will be contacted and offered a position or new advertisement in placed.
R4	The fellow asks for maternity/paternity leave	DC contract will be suspended and extended after the leave; the host country's social security system provides bridging salary.
R5	The fellow requests a part-time research effort to fit in with an external part-time job	It will be explained to the fellow that this is not compatible with MSCA conditions.
R6	The fellow shows repeatedly inappropriate conduct, or misbehaviour	Local disciplinary measures will be adhered to, and LongTREC grievance procedures will be implemented.
R7	DC misses essential training modules	Compensate with local training, using the advice of the TAB and Mentor.
R8	The fellow does not finish the thesis in 36 months	The host will fund the fellow to finish the thesis in a reasonable timeframe.
R9	Conflict between DC and Supervisors	Counselor will mediate. If not successful, the SB will intervene.
R10	The fellow suffers from long-lasting disease	Managed in accordance to each country employment regulations. Research goals will be adjusted and compensated by other DC projects when possible.

Successful key elements

- Clear, appealing and timely research question
- Participants are the EU leaders in the field
- Attractive reading and repetitive message
- Extremely complete training program

Training

Table 1.3b. Main network-wide training event, conferences and beneficiary contributions

#	Joint Event	Lead	Month (days)	SSV Training (see Additional Table)											Other*
															Monitoring
				1	2	3	4	1	2	3	4	1	2	3	
R	**Recruitment workshop**	CSIC	4												R, GA
1	**Welcome retreat**. DCs will meet with the Coordinator and WP leaders to learn about the scope, content, and structure of LongTREC.	CSIC	8 (6)	I II		I			I		I	I	I		TAB
2	**Focus on research**. At LongTREC´s 1-year mark, the DCs will gather with members from SAB in the first full network meeting, allowing DCs to meet *Co-Supervisors* and to choose their future intersectoral *Mentors*.	EI	12 (8)	III	I	II III	I	I		I II					PR, GA
3	**Fellow's retreat**. Science-intensive and soft-skills intensive training. One day will be reserved for community-building among DCs.	IIT	16 (8)	IV V VI	II				II III		II		II		TAB
4	**Mid-term**. The DCs will receive software management training and plan educational materials for Summer School	CNRS	24 (8)	VII VIII	IV	IV V	II	II	VI	III		II		I	PR, GA
5	**Summer school.** DCs disseminate LongTREC by delivering training in lrRNA-seq methods including their newly developed approaches. DCs will create educational materials that will be shared with EU bioinformatics training initiatives such as Elixir, GOBLET, GTPB.	CRG	33 (8)		III	III	III	III	V VI	IV V	III		III		TAB
6	**Conference.** We combine presentation of LongTREC results by DCs with invited talks from external members of the lrRNA-seq community. LongTREC will invite participants to participate in a specialized manuscript Collection at Scientific Data journal. DCs will organize the conference with supervisor support.	CSIC	42 (8)			IV	IV	IV			IV V VI	III IV	IV	II	PR, GA

Detailed description of the training program

Genomics of Gene Expression Lab

Table 1.3b_1 LongTREC Science Training Programme

Third generation, single molecule sequencing technologies

LongTREC DCs will gain understanding in the technical, biochemical and data quality aspects of the different TGS platforms. LongTREC's training in these new sequencing technologies takes the form of seminars about technical and applicability aspects of each different type of LRS method, complemented with hands-on training on most relevant data pre-processing steps to transform raw measurements into useful data. While LongTREC focuses on transcriptomics, talks on other applications will be also included. Learning the basics of LongTREC technologies will be a requisite to all DCs, regardless of the ones used in their projects.

- **Technologies I** (0.5 day): Pacbio and Nanopore technologies (Wobble Genomics[I], Nanopore[I]).
- **Technologies II** (0.5 day): Variety of lrRNA-seq library preparation protocols by LongTREC partners.
- **Technologies III** (1 day). Results of the LRGASP challenge, by Ana Conesa (CSIC) and Angela Brooks (UCSC, LRGASP).
- **Technologies IV** (0.5 day). Genome Assembly and Metagenomics using LRS (speaker T2T Consortium).
- **Technologies V** (0.5 day): Genome Annotation using long reads (Roderic Guigó, CRG, Adam Frankish EBI).
- **Technologies VI** (0.5 day): Single cell applications using long reads (Pascal Barbry, CNRS; Hagen Tilgner, LRGASP).
- **Technologies VII** (0.5 day): Epitranscriptomics using long reads (Eva Maria Novoa; CRG).
- **Technologies VIII** (0.5 day): Long reads proteogenomics: by Gloria Sheynkman (Virginia Medical School; LRGASP).
- **Technologies IX** (0.5 day): lrRNA-seq in non-model organisms by Eduardo Eyras (EMBO Australia, LRGASP participant).
- **Technologies X** (0.25 day): Allele specific expression using long reads: Lauren McIntyre (University of Florida, USA).

Data Analysis, Statistical Science and Software Development

Successful completion of DCs' projects requires expert knowledge of available bioinformatics resources for the analysis of LRS data, as well as knowledge on the statistics of count data and functional genomics. LongTREC DCs will be trained in solid computational biology and biostatistical science to ensure competitiveness and versatility in their career progression. The training program will consist of four workshops, each of them focusing on an important aspect of the knowledge discovery process in computational biology.

🛠 **Tools I. Workshop on the quality control and pre-processing of lrRNA-seq data**. Focus on Data Quality Control. A 4-day workshop that will present current tools and workflows for lrRNA-seq data analysis, in most cases, with training from their creators. The workshop includes generic and LR-specific tools: Isoseq3 (Francisco Pardo, UPV), SQANTI3 (Ana Conesa, CSIC), TALON (Ali Mortazavi, UCI), FLAIR (Angela Brooks, UCSC), TAMA (Richard Kuo[I], Wobble Genomics), TranD (Laure McIntyre, UF), Epinano (Novoa, CRG), uTLRA (Sahlin, SU), Nanocompore (Leonardi, IIt) etc. DCs will work with their data or with LRGASP datasets. They will learn the principles of each tool, analyze data and compare results. Critical discussions will follow to compare their conclusions with that obtained by the LRGASP.

🛠 **Tools II**: **Workshop on Statistical Methods for lrRNA-seq data analysis:** Focus on Statistical Rigor. A 4-day workshop covering the principles of statistical methods used in omics analysis, with a special focus on transcriptomics. Students will learn the data distributions in long and short reads datasets, identification of biases, methods for normalization and for differential expression analysis. The workshop will also include concepts of Multivariate Statistics, Deep learning and AI applied to the omics science. Tools like tappAS, IsoTools and FLAIR, that contain modules for the statistical analysis of long reads data, will be used and students will create their own analysis scripts in R. The workshop will have a hands-on format including group projects. This workshop will be provided by CSIC, CRNS and UF.

🛠 **Tools III: Workshop of Advanced Visualization and curation of lrRNA-seq datasets.** Focus on Biological Conclusions from data. A 4-days workshop that includes visualization and annotation of transcript models in the Genome Browser (Adam Frankish), visualization of epitranscriptomics modifications (Eva Maria Novoa) and Splicing Events (Ali Mortazavi) and functional annotation of alternative isoforms (Ana Conesa). Students working on different projects will present and share their conclusions from data visualization. DCs will write a mini-paper about outcome of training activities from Tools I to III. The last day will be devoted to seminars on the existing research boundaries and challenges in lrRNA-seq. Led by CRG with participation of CSIC, EBI and EI.

🛠 **Tools IV**: **Fundamentals in Software and tool development, and FAIR principles**. Focus on Reproducibility. A 3-days workshop that will cover: good programming skills, programming in group tools, GitHub and Bioconductor, Containers, NextFlow code management, code documentation, creating example data for code, etc. Lectures by software development partners will be included to exemplify how FAIR principles were implemented in their code development. This workshop will be led by BSC (Salvador Capella) who currently leads the Elixir effort in good programming practices and code management.

Translational Science

Translational science is interdisciplinary experimental research that considers practical applications (diagnostics, novel biotech products, therapies) to improve quality of life. This important aspect is hardly covered by many PhD programs, in part because academic research focuses on basic research. Young scientists thus tend to lack significant translational training. Our translational thread is composed of presentations by invited key speakers from industry, SMEs and the clinic. We include seminars from relevant stakeholders that either use omics technologies or include their analysis in their business model. We have the great opportunity of training our DCs in this intersectoral context. Our classes, which will also take the format of a 'fireside chat', will be structured to generate extensive follow-up discussions between instructors, speakers and DCs via online forums, social media and article reports reviewing the latest trends.

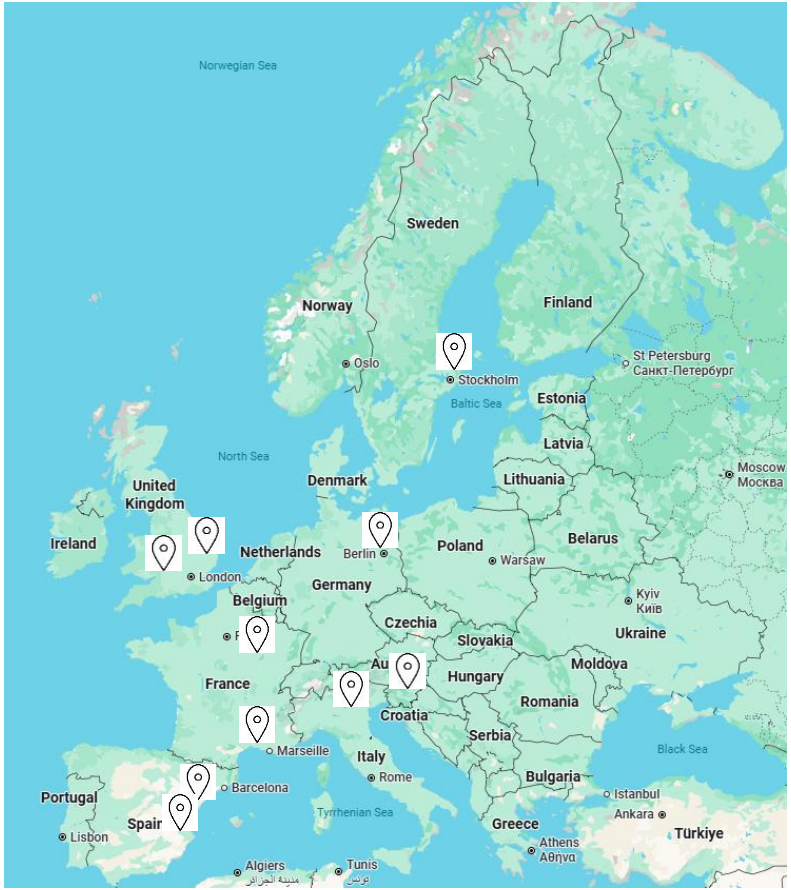
- **Translation I**: Using NGS and omics technologies in the bioindustry (talks by Daniel Ramon[I], ADM and Patricia Diaz[I], IVI)
- **Translation II:** The commercial perspective of third-generation sequencing (talks by Nanopore [I]and Wobble[I])
- **Translation III**: Sequencing methods in the health and pharma industries (Talk by Dr. Fernando Garcia[I], Roche, Strom Therapeutics[I])
- **Translation IV**: From the novel methods to successful commercial bioinformatics tools (Stefan Goetz, Biobam Bioinformatics[I])
- **Translation V**. Commercialization of bioinformatics services (Stefan Goetz, Biobam Bioinformatics[I])

Project Oriented training

The primary focus of each DC should be their research project, as successful completion of the project with significant results is the best way to ensure a good start to the DC's career. However, important aspects for any successful scientist are the identification of bottlenecks during the course of the research, the ability of implementing correction measures and the capacity to be creative beyond the scope of one's own project. To support successful project completion, the TAB and the DCs jointly discuss and review each DCs project yearly. LongTREC will deliver training to stimulate the critical evaluation of own and peer results and the creativity to propose new research ideas.

- **Project I: Explain your project** (1 day): DCs at the beginning of their enrolment will present their project and receive feedback. DCs from different WPs will be paired for peer-to-peer discussions of research proposals.
- **Project II: From Plan to Act** (1 day): Approximately 1 year later, DCs present oral progress updates to LongTREC members. DCs are encouraged to discuss deviations from original plans and contingency measures adopted.
- **Project III: Propose me a grant** (at Summer School): After 18 months of research students will be asked to write a short but fully compliant research grant in the format of a national or international funding agency. During the LongTREC Summer School of LongTREC DCs will present their project idea in the form of an interview with all supervisors participating in the evaluation panel and gather in groups to evaluate project proposals of their peers. This interview follows ERC grant format.
- **Project IV: Show me your results** (at Final Conference): DC project results will be disseminated to the broad audience at the LongTREC

LongTREC DC distribution



- 12 DCs (9 European + 3 non-European)
- 9 nationalities
- Distributed into 7 countries

Recommendations of a less painful management

Genomics of Gene Expression Lab

- No not add too many associated partners. It is a nightmare to collect signatures from the Consortium and YOU will have to chase the people.
- Watch out for big companies. They like to change their minds…
- Make sure the documents reflect what you want, as well as the governance proposed for your project.
- What you include in the proposal, must be implemented…
- Make the start of your project somewhere in March-April to collect master students that are graduating in the summer and can start in September.
- Get a VERY GOOD project manager who is committed with the success of the project and is proactive to find solutions to problems.