



WP4

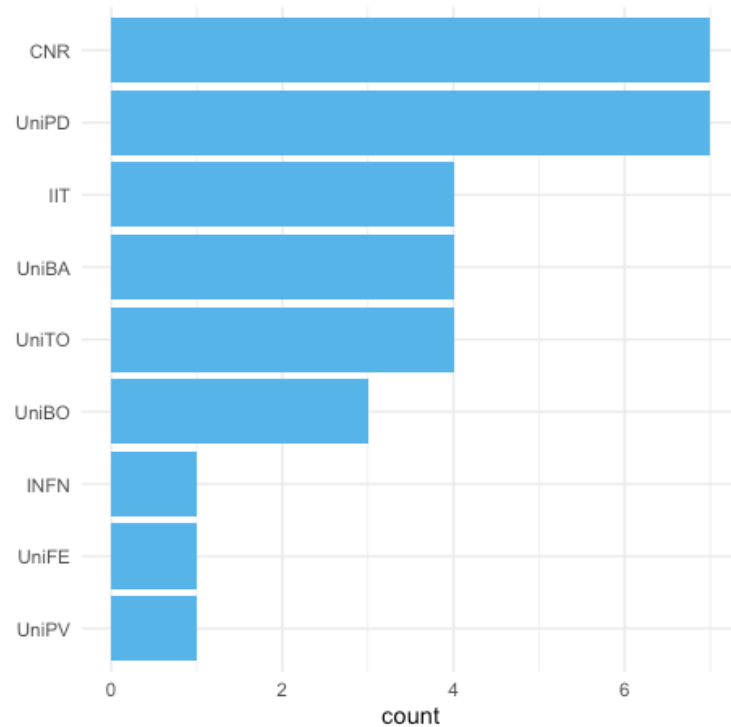
Genome bioinformatic pipelines for GPU-based HPC infrastructures

Chiara Romualdi

Bologna June 21st 2024

Institutions involved

- 7 Universities, CNR and INFN
- 32 affiliates



ISTITUTO ITALIANO DI TECNOLOGIA



WP4 description

Activities:

Development of next-generation bioinformatics tools fully optimized for GPU/FPGA high-performance computing infrastructures to analyse genomic, and transcriptomic data.

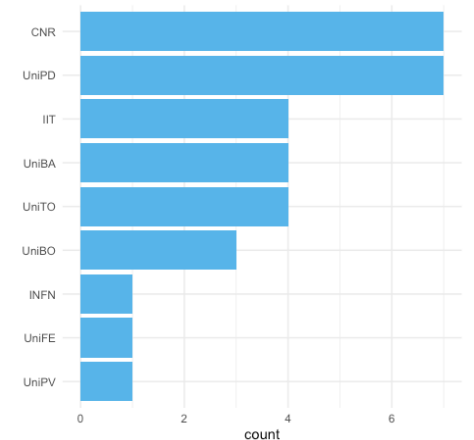


Main Tasks

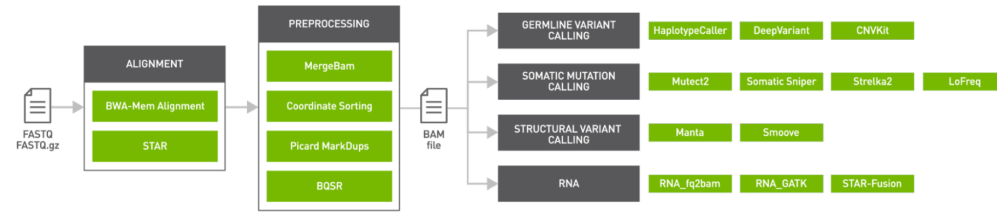
1. **Task 4.1** Benchmark of GPU-based codes for diverse genome sequences
2. **Task 4.2** GPU-based bioinformatics pipelines for single cell transcriptomics
3. **Task 4.3** GPU-based bioinformatics pipelines for long-reads and structural variants



7 Universities, CNR and INFN
32 affiliates



WP4 description



Task 4.1 Benchmark of GPU-based codes for diverse genome sequences

Milestone 4: M1-M18

Referente: Fabio Landuzzi (IIT)

Deliverable: Genome bioinformatic pipelines (M18)

Task 4.2 GPU-based bioinformatics pipelines for single cell transcriptomics

Milestone 4: M9-M36

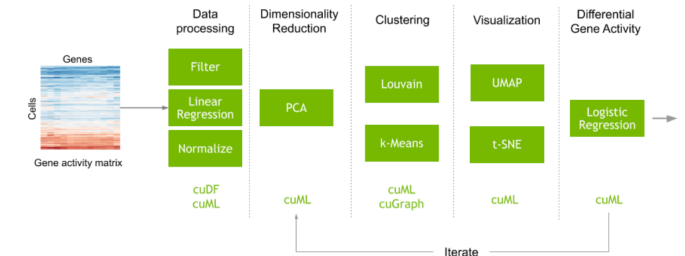
Referente: Raffaele Calogero (UNITO)

Deliverable: Genome bioinformatic pipelines (M36)

NVIDIA-Genomics-Research/**rapids-single-...**



Examples of single-cell genomic analysis accelerated with RAPIDS

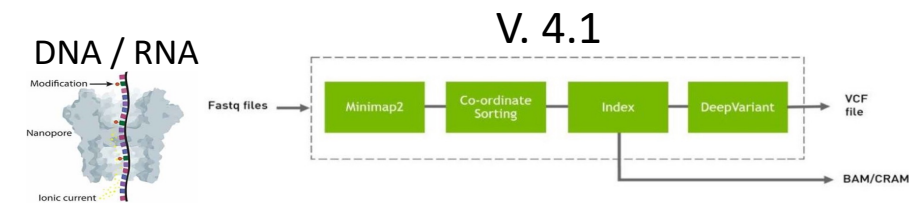


Task 4.3 GPU-based bioinformatics pipelines for long-reads and structural variants

Milestone 4: M13-M36

Referente: Ernesto Picardi (UNIBA)

Deliverable: Genome bioinformatic pipelines (M36)



2. Long-read tooling and workflow available in Parabricks 4.1, with new Minimap2 and FastQ-to-VCF for PacBio

Administrative reports - *recruitment*

1. We have collaboration with **CN3** on methods for circular RNA identification using long reads
2. We have synergies with **WP5** and **WP6**

	# RTDa					# PhD					Other			
	I year	II year	III year	TOTAL		I year	II year	III year	TOTAL		I year	II year	III year	TOTAL
UNIPD	0,5			0,5		2		2				0		
UNITO	1			1				0	1			1		
UNIBA	1			1				0		1		1		
UNIPV	0,3			0,3				0				0		
UNIFE				0	1	1		2		3		3		
UNIBO	0,2			0,2				0				0		
IIT				0		1		1		1		1		
CNR	1	2		3				0				0		
INFN				0				0				0		
TOTAL	4	2	0	6	1	4	0	5	1	5	0	6		

Tutti i reclutamenti sono stati completati

Administrative reports - *organization*

1. Abbiamo pianificato delle riunioni bi-mensili con i referenti dei vari enti affiliati
2. Criteri per la selezione degli interventi nei meeting plenari:
 - *Valutazione degli obiettivi raggiunti*
 - *Fase di avanzamento di un Progetto*
 - *Coerenza con le tematiche dei vari Task*
 - *Turnazione tra gli enti affiliati*
4. Due richieste, per ora *pending* risorse di calcolo al CINECA
 - 1 in fase operative
 - 1 ancora pending ?

Talks summary

Task 2

Cristian Castiglione, UniPD

sgdGMF: a new scalable matrix factorization framework for high-dimensional omics data

Raffaele Calogero, UniTO

Dissecting tumor transcriptional heterogeneity using single-cell RNA-seq data.

Stefano Volinia, UniFE

Integration of scRNAseq datasets for studying cellular heterogeneity in solid tumors.

Task 3

Ernesto Picardi, UniBA

Inosinome profiling by high-throughput sequencing