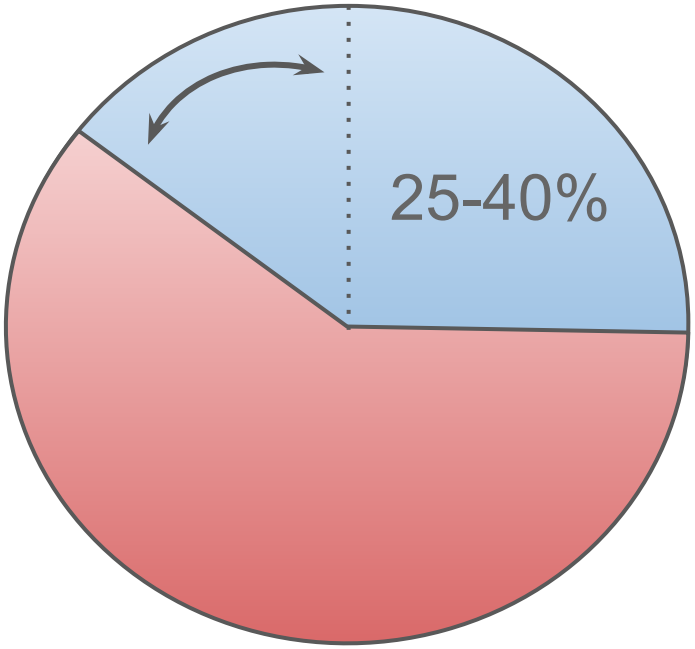
Remus - a web-based tool for searching pathogenic variants in tissue-specific regulatory regions

Paweł Sztromwasser
BTM, Medical University of Lodz

NATIONAL SCIENCE CENTRE
POLAND

Diagnostic yield



WES, gene panels

Gilissen et al. 2014; Yang et al. 2013; Lee et al. 2015; Valencia et al. 2015

Where are the missing diagnoses?

1. Not a monogenic disease

Where are the missing diagnoses?

1. Not a monogenic disease
2. Structural variants

Where are the missing diagnoses?

1. Not a monogenic disease
2. Structural variants
3. Limitations of short reads (~84% of the genome is mappable)

Where are the missing diagnoses?

1. Not a monogenic disease
2. Structural variants
3. Limitations of short reads (~84% of the genome is mappable)
4. Incomplete reference genome (we know ~94% of it)

Where are the missing diagnoses?

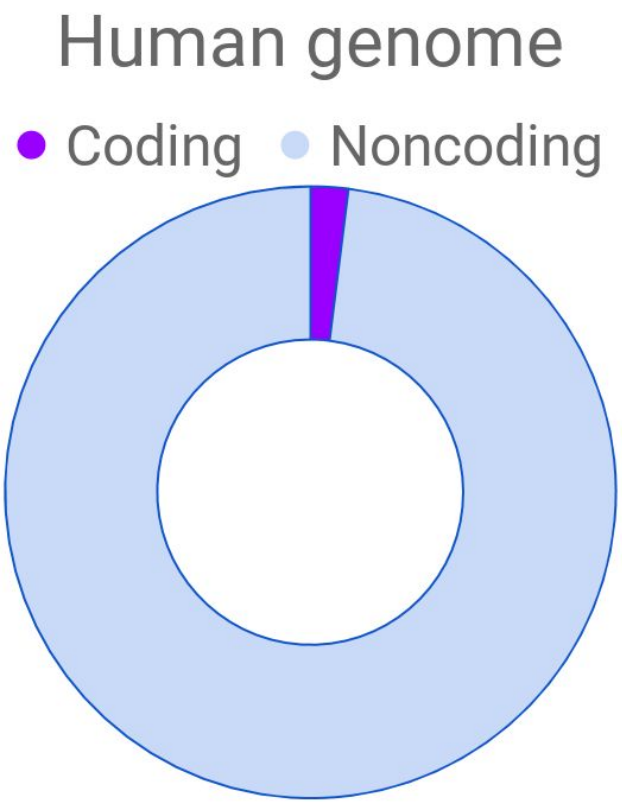
1. Not a monogenic disease
2. Structural variants
3. Limitations of short reads (~84% of the genome is mappable)
4. Incomplete reference genome (we know ~94% of it)
5. Imperfect bioinformatic tools / algorithms

Where are the missing diagnoses?

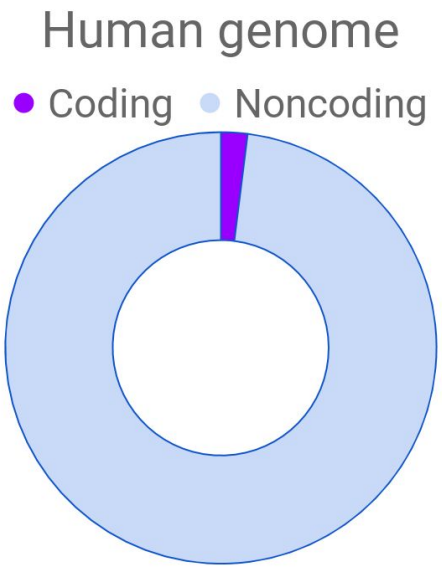
1. Not a monogenic disease
2. Structural variants
3. Limitations of short reads (~84% of the genome is mappable)
4. Incomplete reference genome (we know ~94% of it)
5. Imperfect bioinformatic tools / algorithms

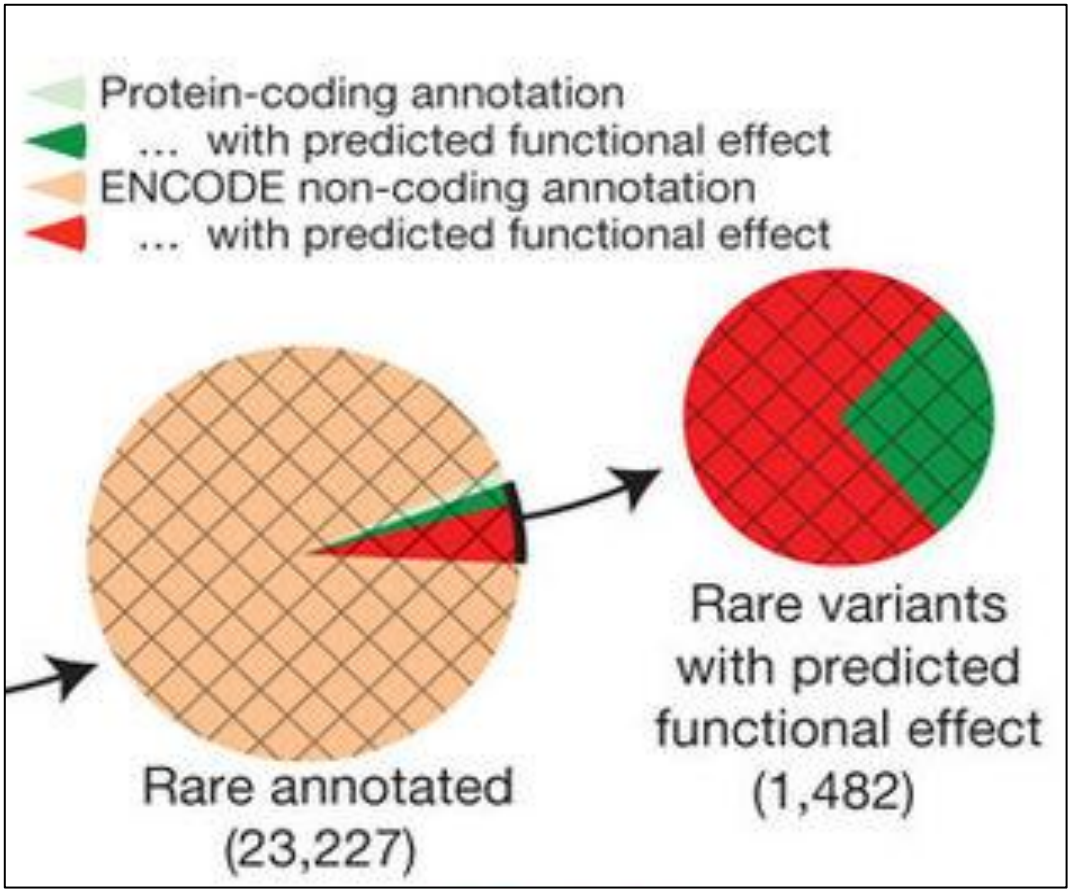
Genes-only analysis!

Gene based analysis



Gene based analysis





The ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome, Nature (2012).

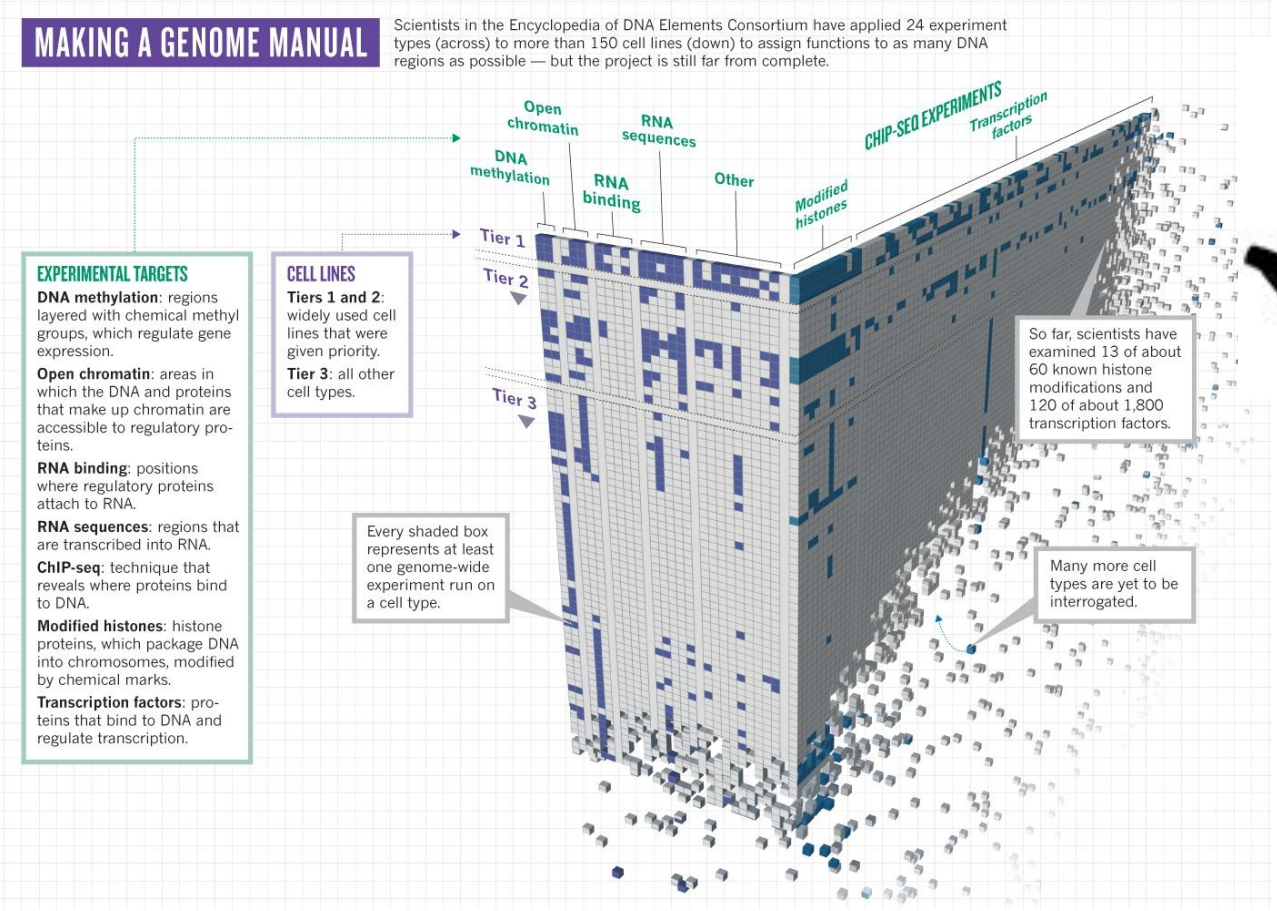
453

cases of **noncoding**, **regulatory** mutations causing a rare monogenic disease, according to **OMIM** database

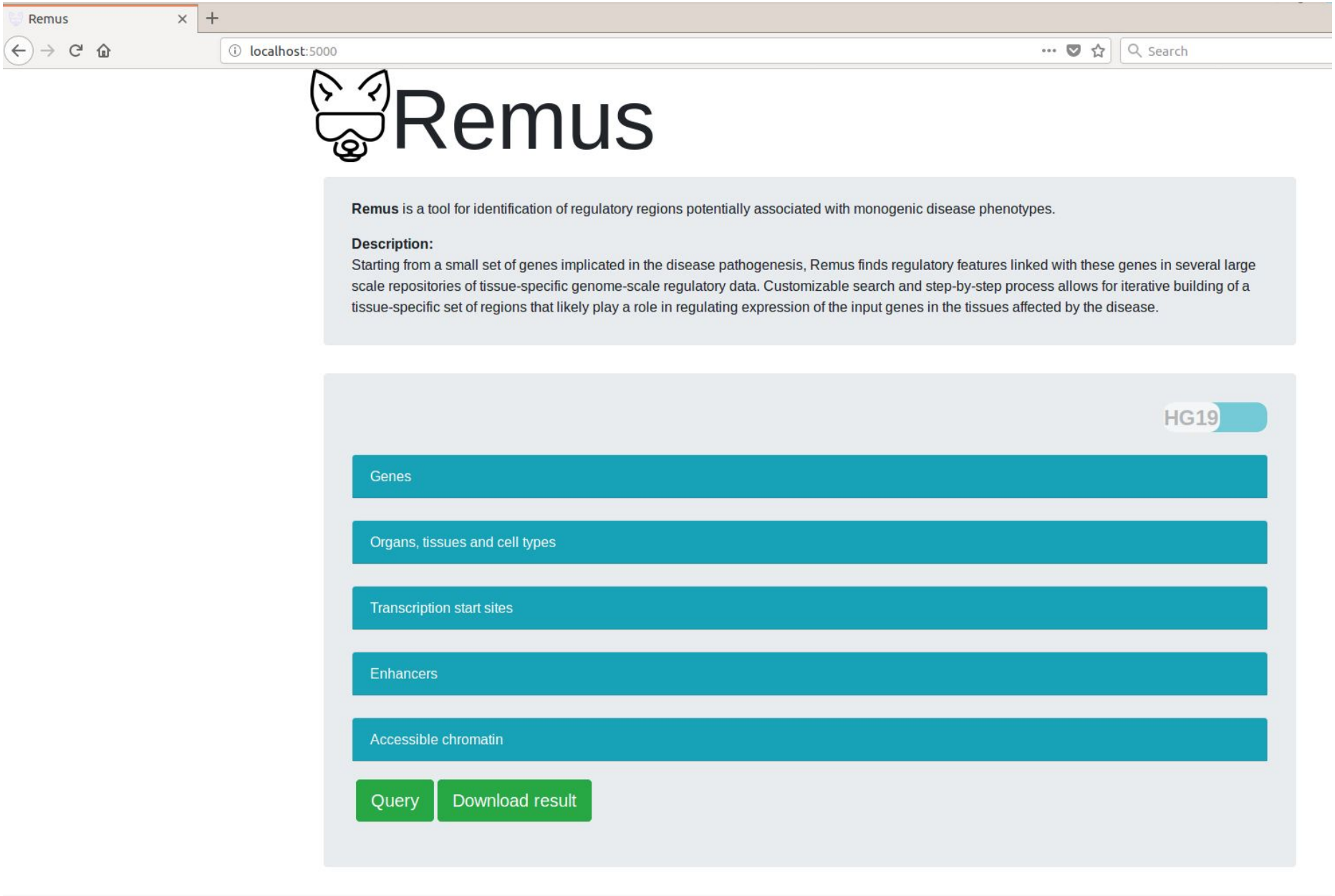
Smedley D. et al., A Whole-Genome Analysis Framework for Effective Identification of Pathogenic Regulatory Variants in Mendelian Disease, The American Journal of Human Genetics (2016)

Why are regulatory variants ignored?

Regulatory DNA



Maher, B. ENCODE: The human encyclopaedia. Nature (2012)


Remus
localhost:5000
Search
Remus
Remus is a tool for identification of regulatory regions potentially associated with monogenic disease phenotypes.
Description:
Starting from a small set of genes implicated in the disease pathogenesis, Remus finds regulatory features linked with these genes in several large scale repositories of tissue-specific genome-scale regulatory data. Customizable search and step-by-step process allows for iterative building of a tissue-specific set of regions that likely play a role in regulating expression of the input genes in the tissues affected by the disease.
HG19
Genes
Organs, tissues and cell types
Transcription start sites
Enhancers
Accessible chromatin
Query
Download result

BTM

MEDICAL UNIVERSITY OF LODZ

HG19

Genes

Select genes:

×HNF1B HNF

HNF1A

HNF1A-AS1

HNF1B

HNF4A

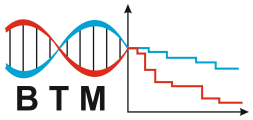
HNF4G

Transcription start sites

Enhancers

Accessible chromatin

Query | Download result



MEDICAL UNIVERSITY OF LODZ

HG19

Genes

Select genes:

×HNF1B

[Tips]

Organs, tissues and cell types

Select organs/tissues/cell types:

×kidney epithelial cell (ENH_F5, CHRM)
×kidney other (ENH_F5, CHRM) li

iris pigment epithelial cell (ENH_F5, CHRM)
kidney epithelial cell (ENH_F5, CHRM)
lens epithelial cell (ENH_F5)
liver other (ENH_F5, CHRM)
lung microvascular endothelial cell (CHRM)
lymphocyte of B lineage (ENH_F5)

Enhancers

Accessible chromatin

Query | Download result

Organs, tissues and cell types

Select organs/tissues/cell types:

×kidney epithelial cell (ENH_F5, CHRM)
×kidney other (ENH_F5, CHRM) ×liver other (ENH_F5, CHRM)

[Meaning of symbols in parentheses]

[Tips]

Transcription start sites

☑ FANTOM5

- ○ Active in all selected tissues [?]
- ○ Active in any of selected tissues [?]

bps upstream

3000

bps downstream

100

Enhancers

☑ FANTOM5

- ◉ Active in all selected tissues
- ○ Active in any of selected tissues

Kb upstream

500

Kb downstream

500

☐ ENCODE

Transcription start sites

FANTOM5

Active in all selected tissues [?]

Active in any of selected tissues [?]

bps upstream

3000

bps downstream

100

Enhancers

FANTOM5

Active in all selected tissues

Active in any of selected tissues

Kb upstream

500

Kb downstream

500

ENCODE

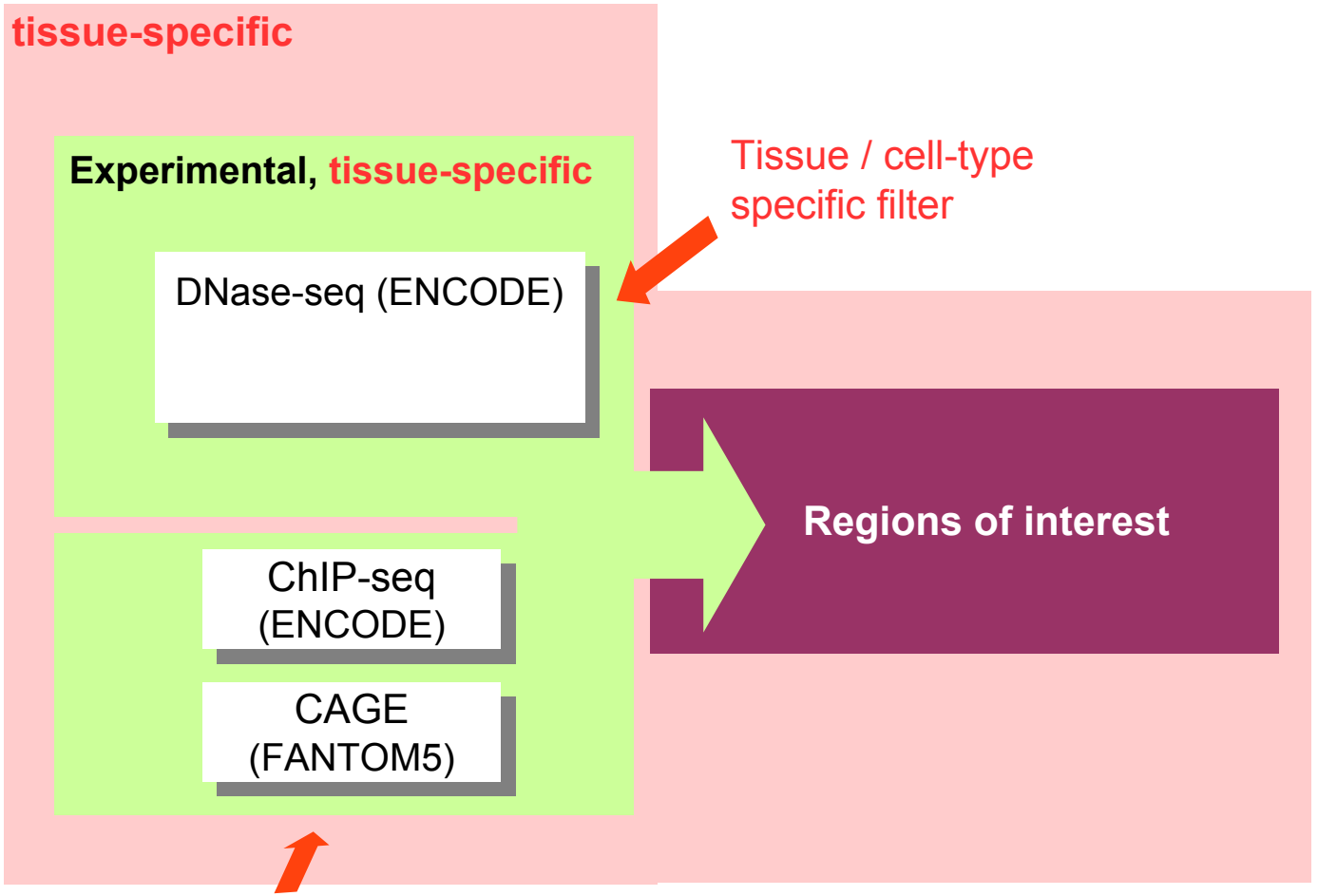
Accessible chromatin

Query | Download result

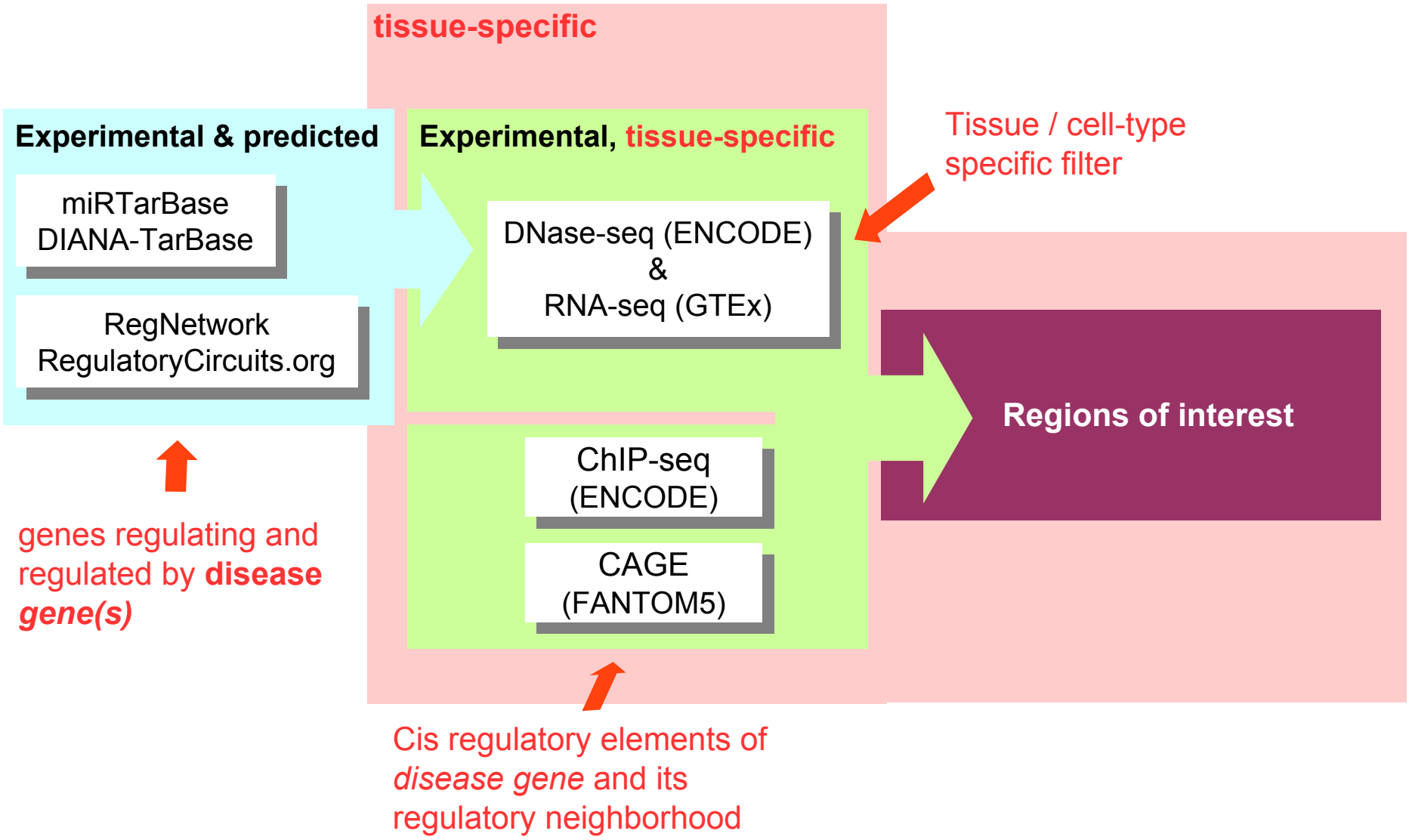
Summary table

Time elapsed (s)	0.950094
No. features	7
No. base pairs	60226

Remus under the hood



Remus under the hood



Availability

https://github.com/DamianSkrzypczak/Remus

https://biostat.umed.pl/polonez_eng.html

Acknowledgements

Damian Skrzypczak

Wojciech Fendler

https://github.com/DamianSkrzypczak/Remus

https://biostat.umed.pl/polonez_eng.html

Projekt finansowany przez Narodowe Centrum Nauki poprzez granty:
OPUS 2014/15/B/NZ5/00144 oraz POLONEZ 2016/23/P/NZ2/04251.

Projekt finansowany ze środków przeznaczonych na program finansowania badań naukowych i innowacji UE "Horyzont 2020" na podstawie umowy Nr 665778 o dofinansowanie działań "Marie Skłodowska-Curie.

Diagnostics of rare genetic disorders

~7000 disorders affecting ~7% of population (30mln in Europe)

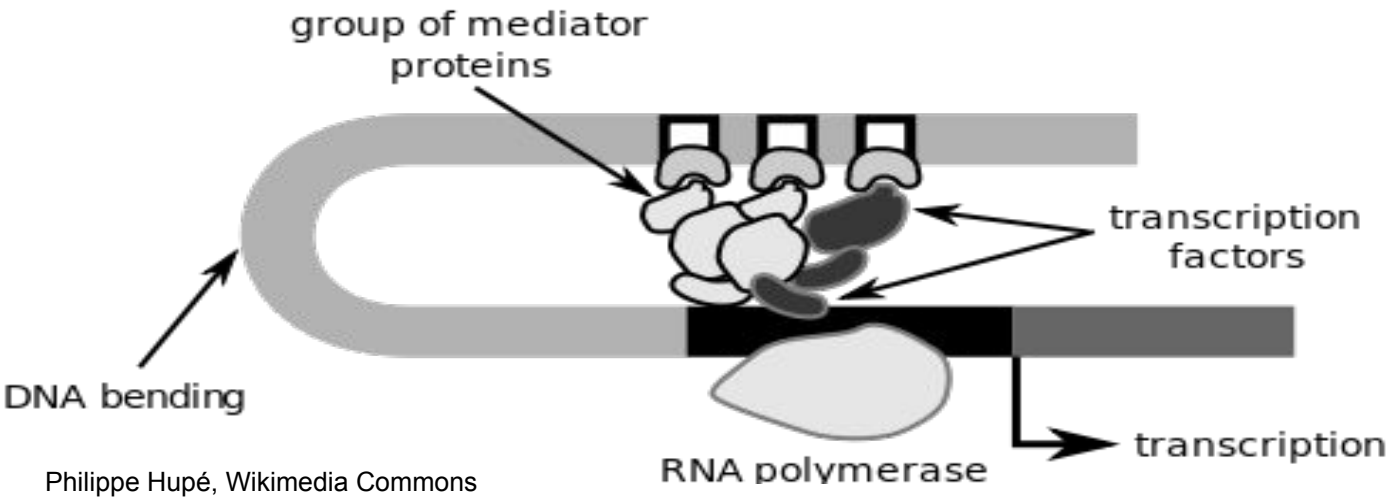
75% affects children under 2yrs

In 80% cases it leads to premature death

~3,000 - 4,000 are monogenic

Diagnosis can determine treatment

Exemplary non-coding regions



Philippe Hupé, Wikimedia Commons