Life Sciences, Physiology and Medicine

Voice for Purpose phase 2 HPC status report

HPC User Day
October 1st, 2025

Fabio Minazzi
Director of Audiovisual

translated.

Translated profile

- **Main services**
  - **Language services** (Translation, Transcreation, Subtitling, Dubbing)
  - **Human-AI** Symbiosis in Translation

- **AI experience (products, services, research):**
  - **ModernMT/Lara:** Machine translation/LLM solutions
  - **Matecat:** Computer-assisted translation system
  - **Matesub / Matedub:** AI-based Subtitling / Dubbing
  - Meetween: multimodal real time video communication translation
  - DVPs: multimodal foundation model for communication
  - First Application Transformer 2017
  - Largest translation dataset and highest quality in the world
  - Best FP7 and H2020 Language AI Research in Europe
    - Matecat - ModernMT (2010 / 2018)

Voice for Purpose: Giving voice to those who cannot speak

- **Give expressive voices to people affected by:**
  Motor neuron conditions (ALS, SMA), Lanynx Cancer, Stroke, Cerebral Palsy, Autism.

- **Main project features**
  - User profiling and medical assistance
  - Voice Donation, Voice Self-Donation
  - Expressive AI voice model creation
  - Real time delivery of expressive AI speech to users
- **Innovations**
  - **Voice donation:** social platform to support people in need of voice
  - **Cloud-based:** infrastructure for delivery to any device
  - **Expressive AI voices:** high quality, real time voice rendering

- **Large social footprint:** 5000+ donors, 100+ users
- **Findings:** measured users preference for self-donated or donated voices over standard impersonal voices

Patient Testimonial

Voice for Purpose: Testimonial

The Voice for Purpose project - Phase 2

Research Problems:

1. Improve input speed/fatigue for people using AAC with eye-gaze

} => • **LLM track**

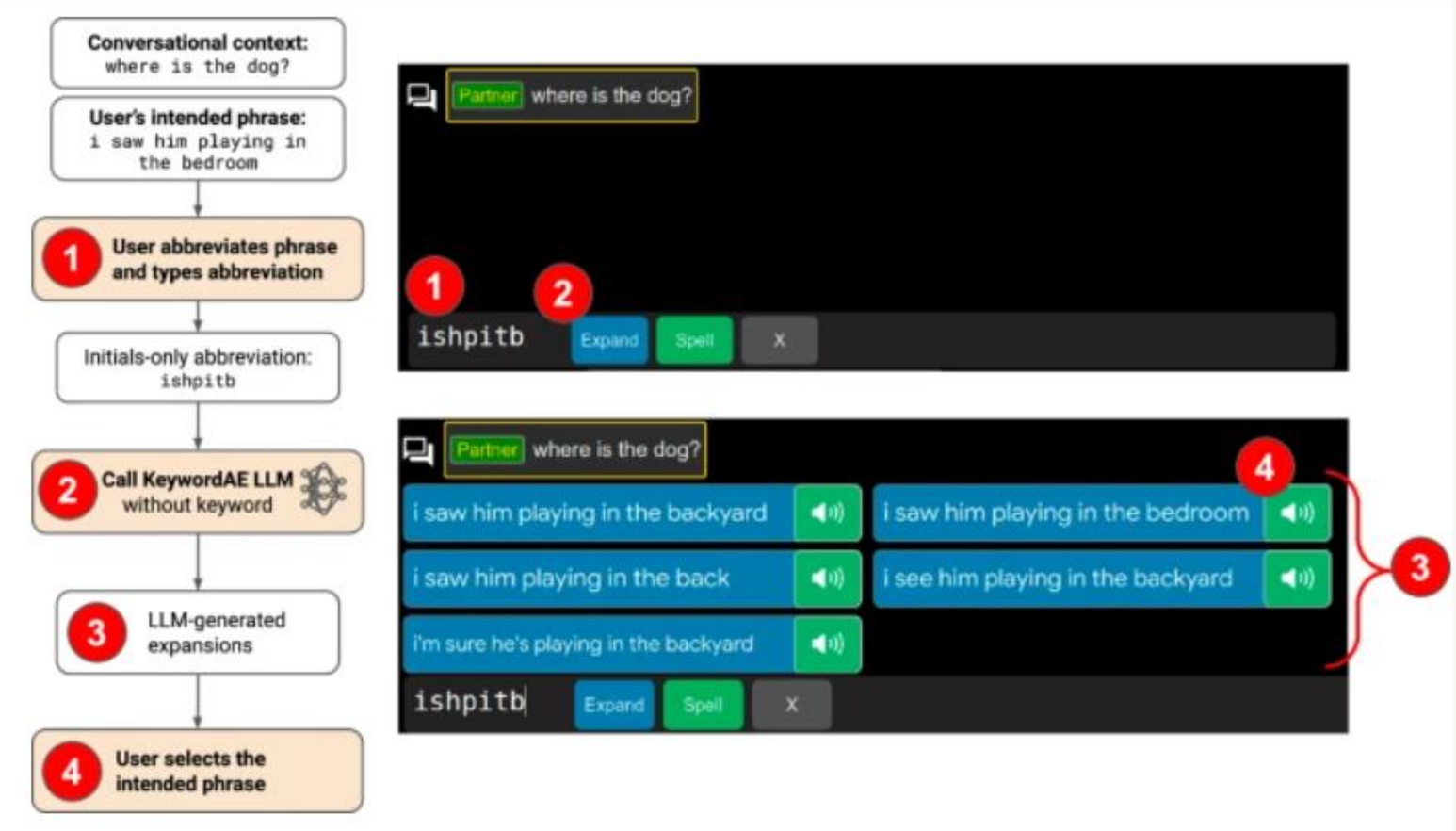
1. Reduce the dataset recording for personal voices
2. Improve models prosody
3. Improve multilingual performance
4. Reduce number of voice models to maintain

} => • **Audio track**

LLM track

Task Reproduce Speakfaster paper by Google, and replicate it in multiple languages.

Concept: Predict sentences from word initials, minimizing number of clicks cognitive effort => increase speed



LLM track

Models tested for prediction:

Open source

- Llama 8B, base and instruct
- Qwen 1B, 3B, 7B, base and instruct
- Microsoft Phi-3-small 7B
- OpenHermes 2.5 Mistral 7B

Close source

- GPT 4.1 and 4.1 mini

Tested: temperature manipulation, few shot, prompt engineering, and oversampling.

Current results: Llama 8B base doubles accuracy over baseline (10% => 20%)

Ongoing: Finetune Llama 8B and 70B base on LEONARDO.

Expected accuracy improvement: 300% compared with non finetuned (20% => 60% accuracy)

Expected Keystrokes reduction: 63%, lowering patient fatigue (SOTA reduces by 48%)

Speed improvement goal: Goal: 30%-50%

TTS track

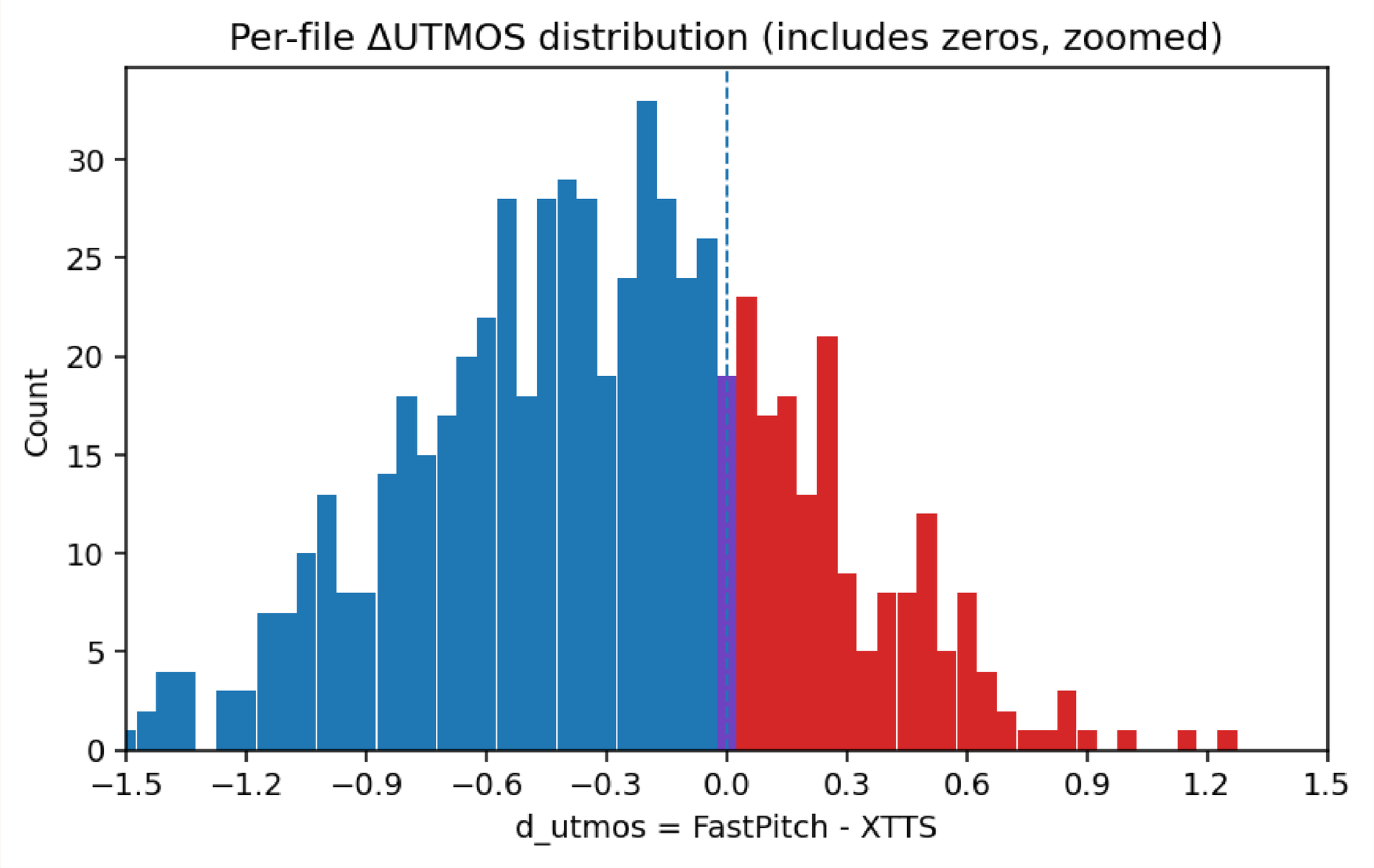
- **Baseline:** Fastpitch (0.08b) finetuned: 150h of voice datasets and 2h of specific users' voices
- **New model:** XTTS v2 (0.5B) finetuned: 150h of voice datasets. Condit. 2min of specific user's voice
- **Tested on 34 speakers, incl. professional voice actors, healthy people, voice impaired people**

Results

- WER (Whisper): XTTS: WER 1.5% Fastpitch 2.4%
- UTMOS: XTTS vs Fastpitch: +0.5 pts avg
- **XTTS produces more fluent, multilingual voices, with less data**

Male-Italian

Female-Spanish



XTTS v2 rates better than Fastpitch on the 1-5 UTMOS scale

Fastpitch rates better than XTTS v2 on the 1-5 UTMOS scale

Computing Resource usage

AI and Data Intensive Applications Access call
Awarded 200k hours on LEONARDO (Cineca, IT): study new, large voice and language models models.

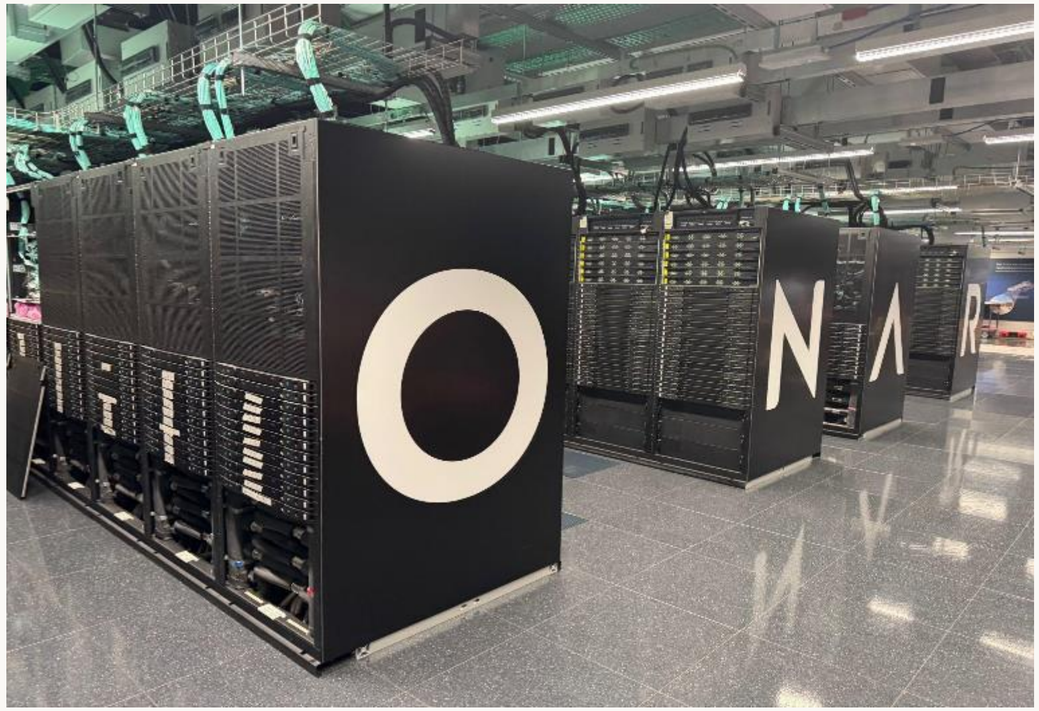
- **Llama 8B**
  - single training 32 gpus 1 day: 800 h
  - Hyperparameter optimisations (guess): **2,500 h**
- **Llama 70B**
  - single training 64 gpus 5 days: 8,040 h
  - 10 languages: 8,040 * 10 = **80,400 h**
  - Hyperparameter optimisations (guess): **10,000 h**
- **total LLM track = 92,900 h**

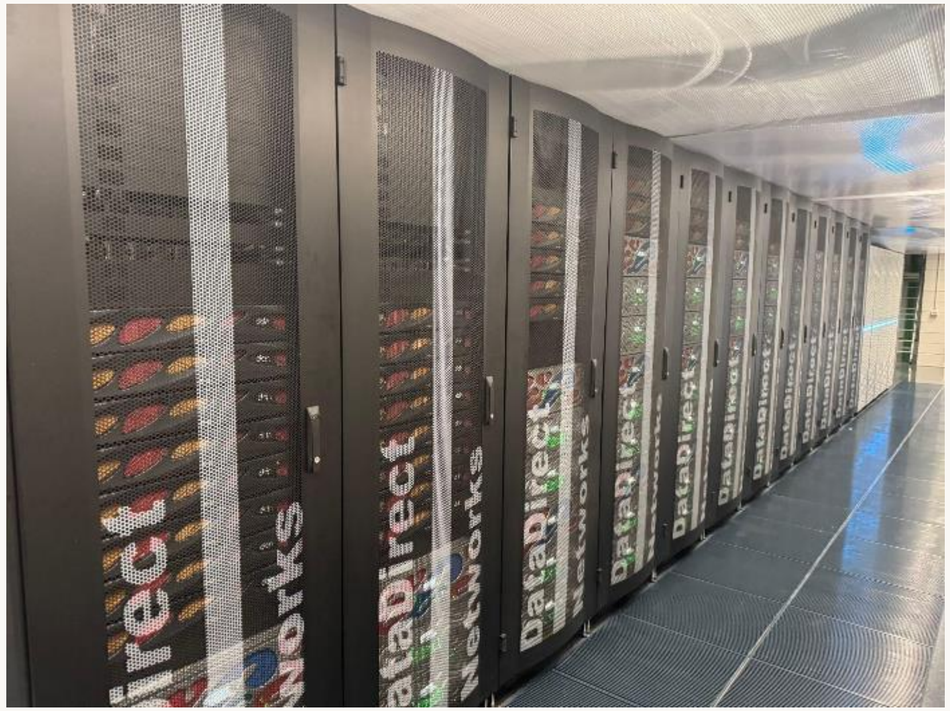
- **XTTS:**
  - Single fine-tuning: 24h on 4 GPUs (per encoder and decoder), i.e. 192h
  - 34 trainings resulting in: 34 * 192 h = **6528 h**
  - Hyperparameter optimisations: **5000 h**
- **KoelTTS**
  - Single training: 32 GPUs for 5 days: 3,840 h
  - 10 language specific models: 10 * 3,840 h = **38,400 h**
  - One big model covering all languages: 64 GPUs for 7 days: **10,752 h**
  - Hyperparameter optimisations: **25,000 h**
- **Total TTS track: 85,680 h**

Conclusions

HPC very powerful infrastructure to:

- run training of large scale models
- implement flexible research practices in different fields





Thank you

Any questions?
We’re here for you.

Fabio Minazzi
Director of Audiovisual
fabio@translated.com

translated.

ModTox: A Platform for toxicity predictions from simulations of high risk off - target proteins

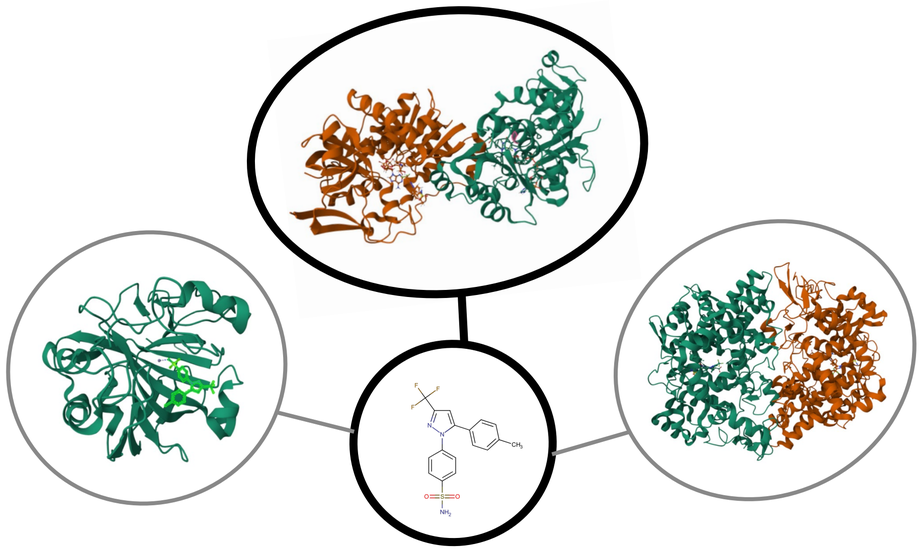
Outline

- Model for small molecule toxicity
- High risk off-target proteins
- MD simulations
- Database creation
- Data analysis

Model for small molecule toxicity

Toxicity ~ f(potency, promiscuity) + other pathways

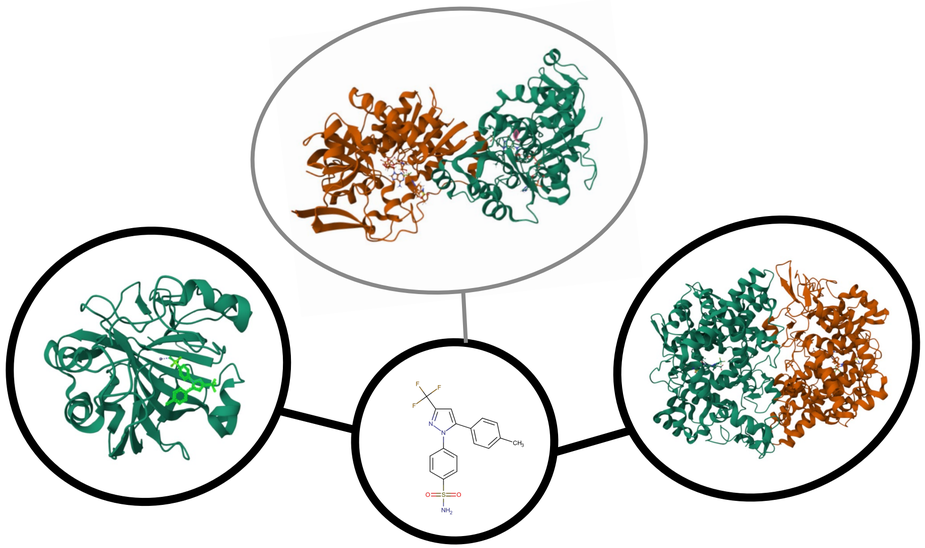
average on - target
binding affinity



Model for small molecule toxicity

average off - target
binding affinity

Toxicity ~ f(potency, promiscuity) + other pathways

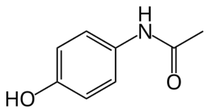


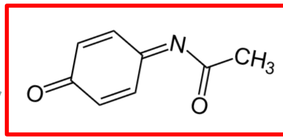
Model for small molecule toxicity

Toxicity ~ f(potency, promiscuity) + other pathways

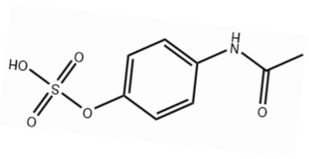
Acetaminophen

(Paracetamol)

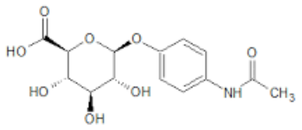




N-acetyl p-benzoquinone



Acetaminophen sulfate



Acetaminophen glucoronide

undesired on - target effect metabolites

…

Model for small molecule toxicity

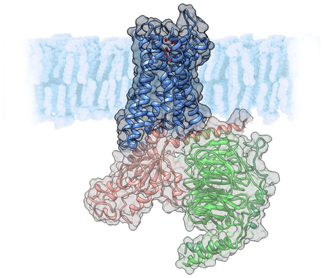
Toxicity ~ f(potency, promiscuity) + other pathways

Maximizing potency
+
Minimizing promiscuity

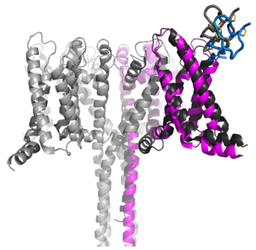
→ Increased selectivity → Lower risk of off-target interactions → Reduced risk of side-effects

Identification of high risk off - targets

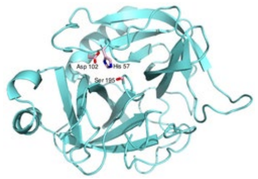
- **Minimal panel from industry** : set of unintended biological targets that pharma companies use for early safety screening (AstraZeneca, GlaxoSmithKline, Novartis and Pfizer)
- **High risk off - targets (HROTs) frequently limiting maximum recommended therapeutic dose** : Found high risk off - targets identifying the proteins that bound most often to high promiscuous, low dose drugs.



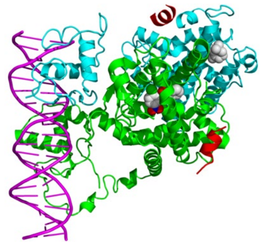
GPCRs



Ion channels



Enzymes (kinases)

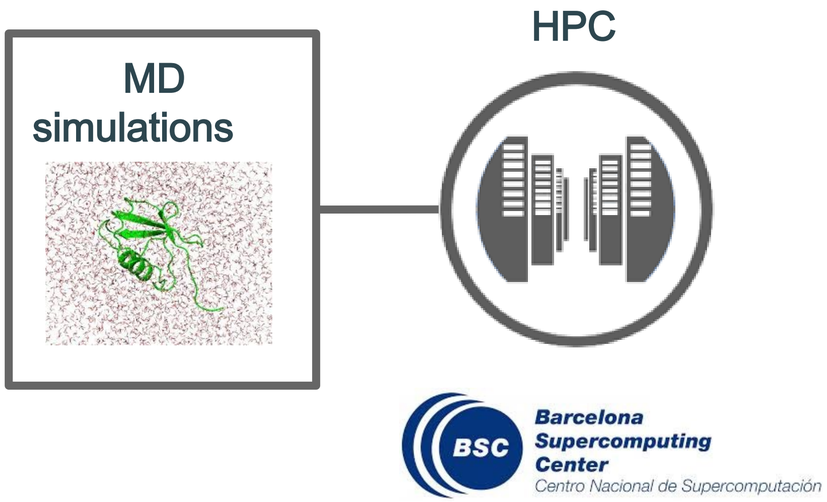


Nuclear receptors

...

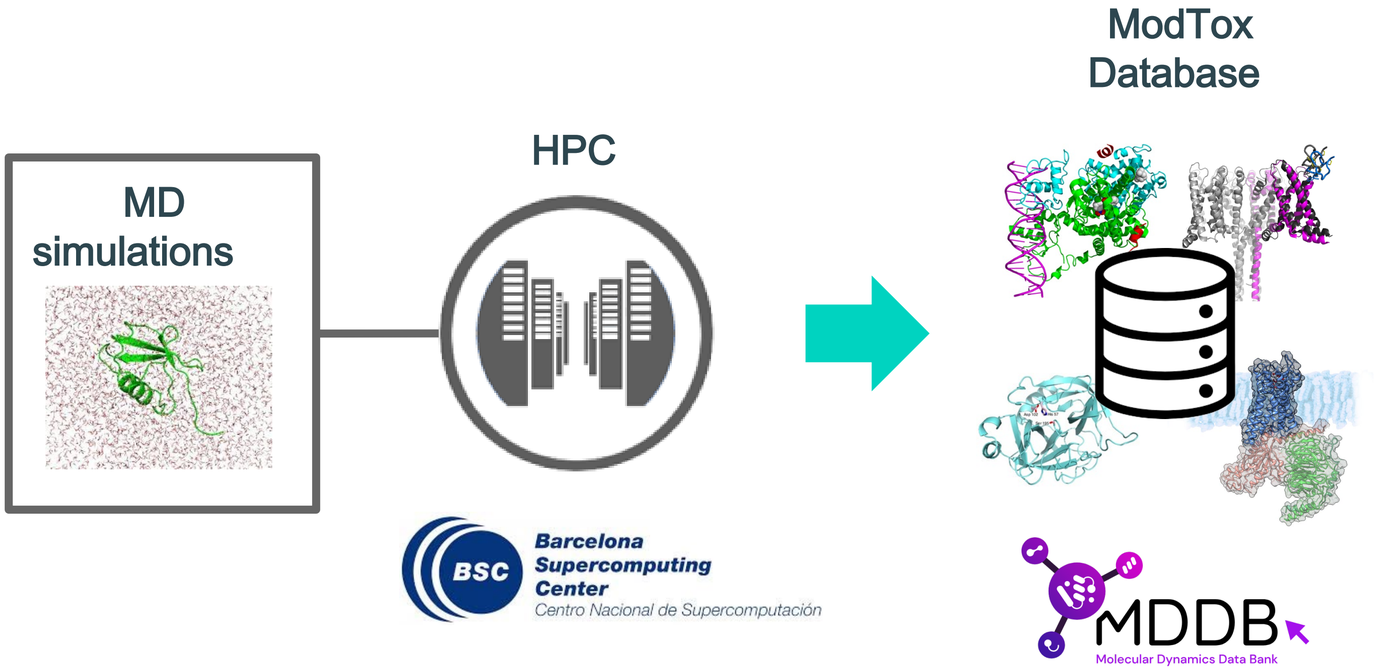
Estimation of Maximum Recommended Therapeutic Dose Using Predicted Promiscuity and Potency. *Clinical Translational Sci* 2016
Reducing Safety-Related Drug Attrition: The Use of in Vitro Pharmacological Profiling. *Nat Rev Drug Discov* 2012

ModTox database and platform



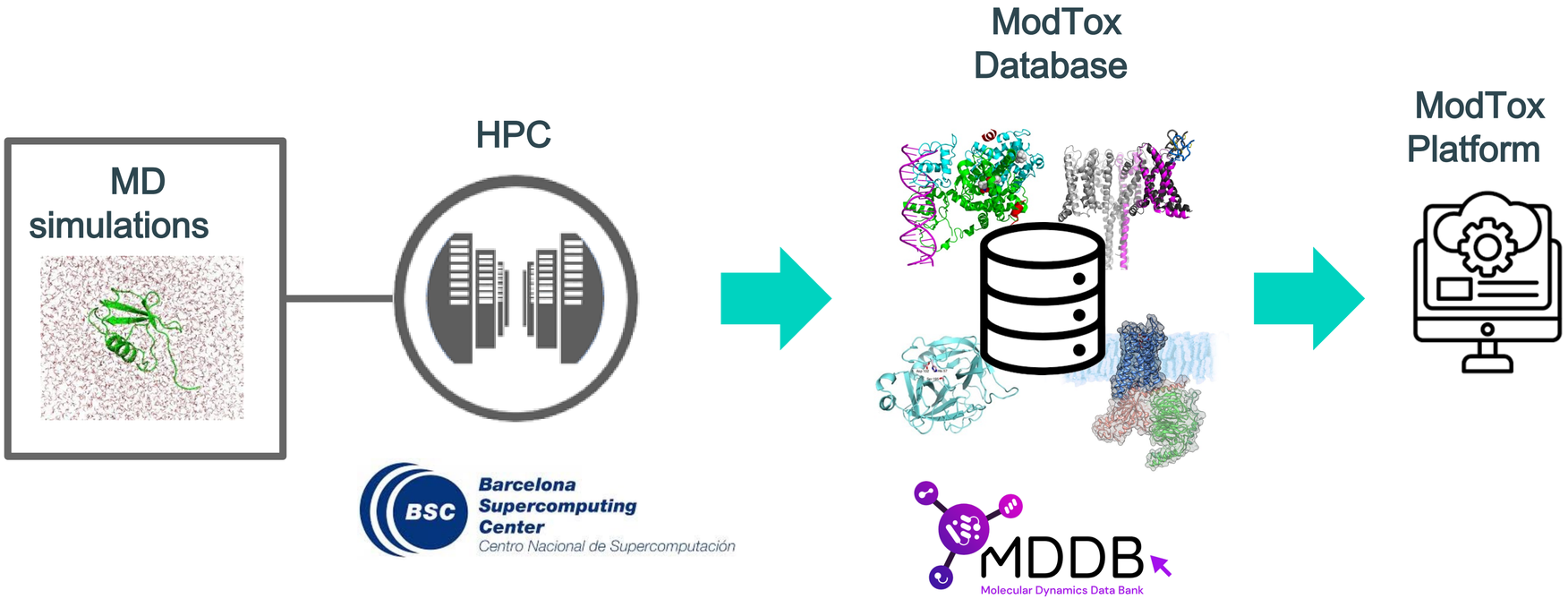
- Generate **conformational ensembles** through all - atom MD simulations of off - target proteins

ModTox database and platform



- Generate **conformational ensembles** through all - atom MD simulations of off - target proteins
- Connect ModTox to the **Molecular Dynamics DataBase** (MDDB)

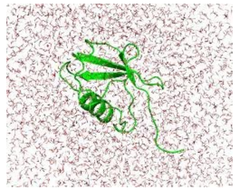
ModTox database and platform



- Generate **conformational ensembles** through all - atom MD simulations of off - target proteins
- Connect ModTox to the **Molecular Dynamics DataBase** (MDDB)
- Create analysis and **models** to assess **toxicity** of new compounds

MD simulations: BioBB pipeline

MD simulations



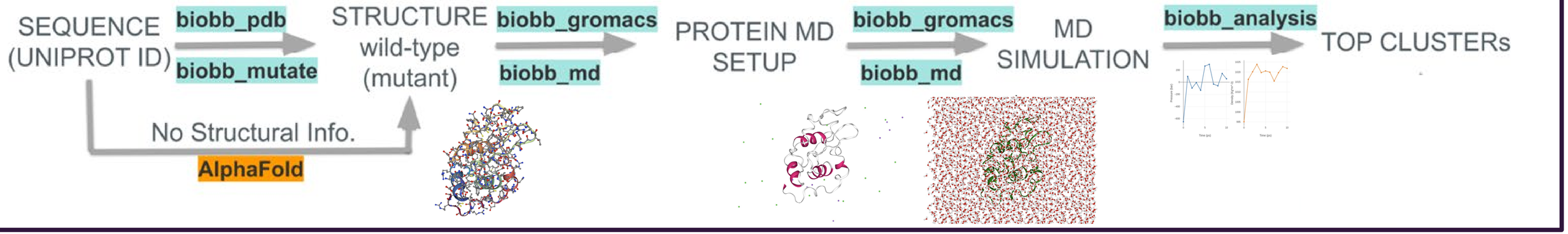
- **Automation** achieved through **BioBBs** - python library to build biomolecular simulation workflows

1. Obtain initial structure → PDB or AF2
2. Fix defects
3. Prepare simulation
4. Minimization, equilibration and production → 3 replicas x 500 ns
5. Post-processing and clustering

GROMACS
FAST. FLEXIBLE. FREE.

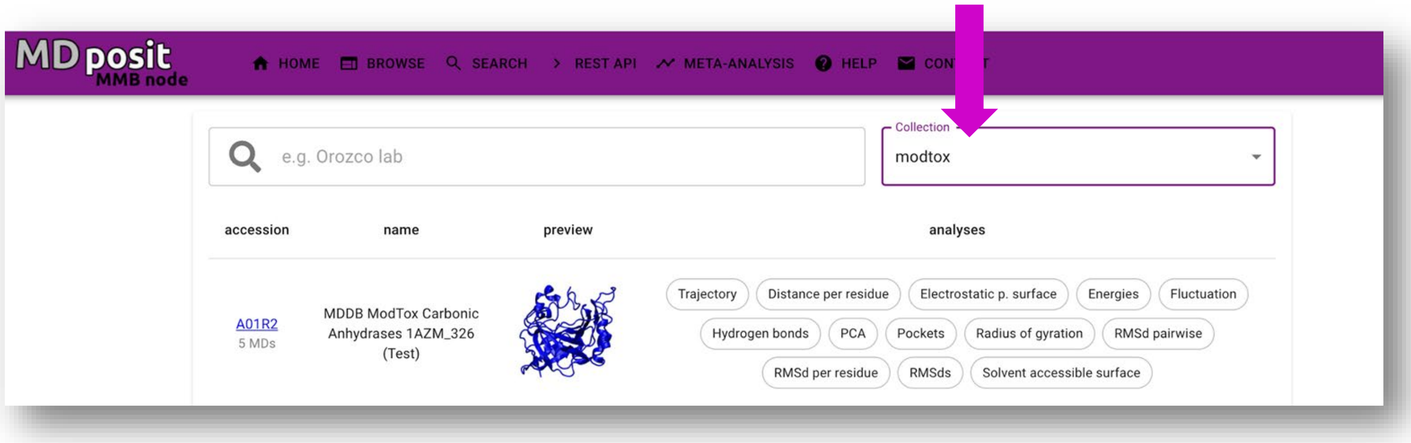
biobb

Biobb Workflow



ModTox database in MDDB

- MDDB is a European initiative to build a federated database of MD simulations
- The trajectory is loaded into a local database, then connected to a global server
- Simple analysis are readily available



Quality control

- RMSDs
- RMSD per residue
- RMSD pairwise
- Radius of gyration
- Fluctuation
- PCA
- Solvent accessible surface
- Clusters
- Dihedral energies

Interactions

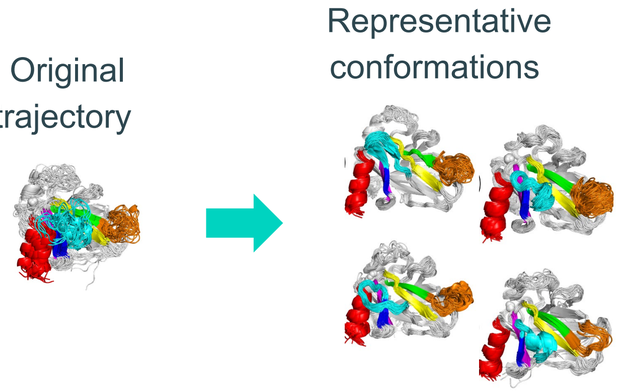
- Distance per residue
- Electrostatic p. surface
- Hydrogen bonds
- Energies

Other

- Pockets

Analysis and models

- Extract representative conformations from binding sites

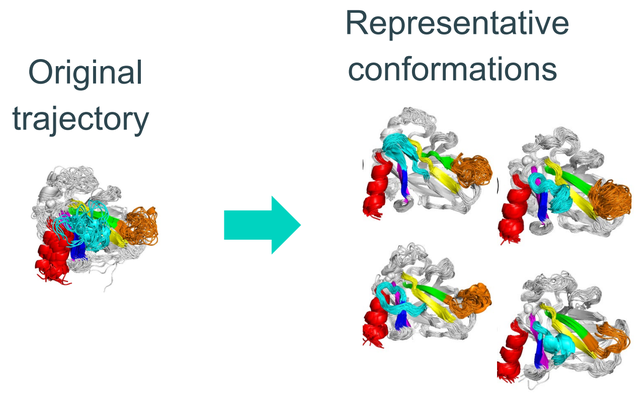


- Test toxicity model on a set of active and decoy ligands

Toxicity ~ f(potency, promiscuity) + other pathways

Analysis and models

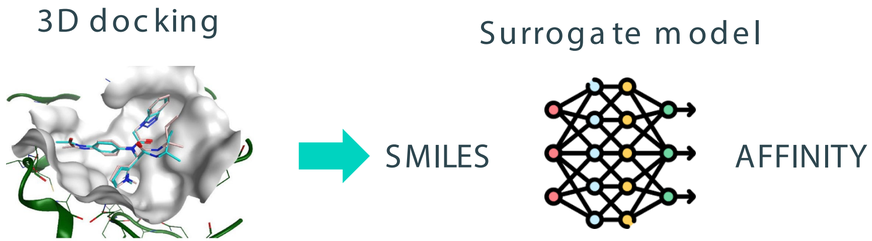
- Extract representative conformations from binding sites



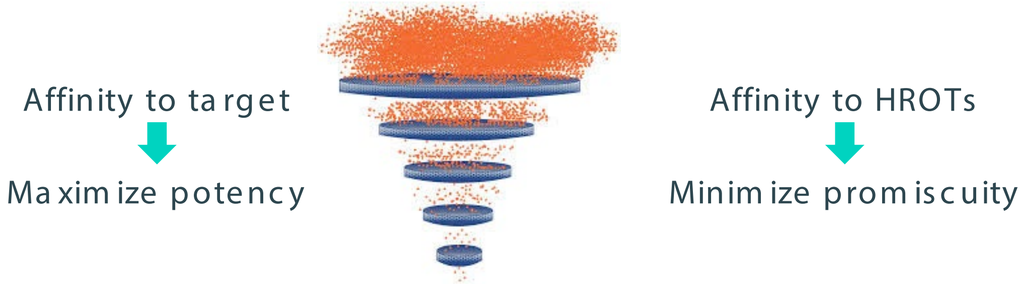
- Test toxicity model on a set of active and decoy ligands

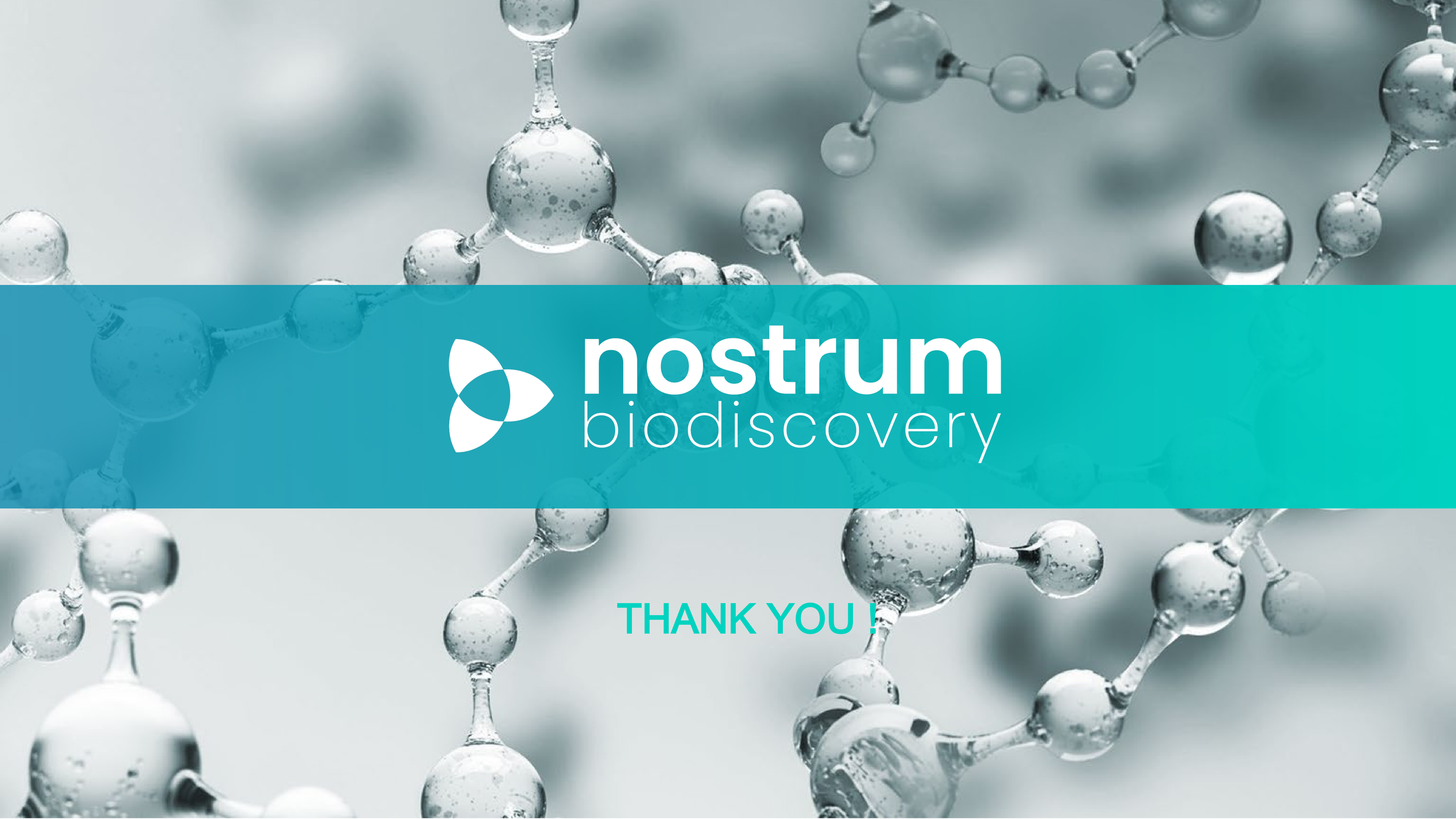
Toxicity ~ f(potency, promiscuity) + other pathways

- Train surrogate models to approximate affinity to HROTs



- Use DSD models to screen compounds during HTVS





nostrum
biodiscovery

THANK YOU !

References

1. Bowes, J.; Brown, A. J.; Hamon, J.; Jarolimek, W.; Sridhar, A.; Waldron, G.; Whitebread, S. Reducing Safety-Related Drug Attrition: The Use of in Vitro Pharmacological Profiling. *Nat Rev Drug Discov* 2012, *11* (12), 909–922
2. Liu, T.; Altman, R. B. Relating Essential Proteins to Drug Side-Effects Using Canonical Component Analysis: A Structure-Based Approach. *J. Chem. Inf. Model.* 2015, *55* (7), 1483–1494.
3. Liu, T.; Oprea, T.; Ursu, O.; Hasselgren, C.; Altman, R. Estimation of Maximum Recommended Therapeutic Dose Using Predicted Promiscuity and Potency. *Clinical Translational Sci* 2016, *9* (6), 311–320

Porting Epistasis Detection Methods to EuroHPC Supercomputers

Ricardo Nobre, Aleksandar Ilic, and Leonel Sousa

INESC-ID, Instituto Superior Técnico, Universidade de Lisboa, Portugal

{ricardo.nobre,aleksandar.ilic,leonel.sousa}@inesc-id.pt

EuroHPC User Days 2025
September 30 - October 1, 2025 in Copenhagen, Denmark

SNP association studies and applications

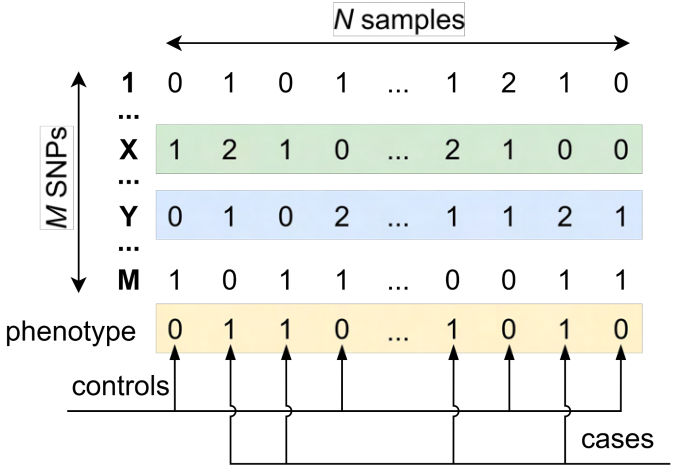
Search for statistical associations between genetic markers and a given trait

Processing over case-control datasets

case: individual with the trait under study
control: individual that does not have the trait

Useful for a number of use cases

Personalized medicine, drug development, mitigate the spread of viruses, forensics



Datasets represent bi-allelic SNPs

SNP: Single nucleotide polymorphism
Major allele: most frequent, **minor allele**: least frequent

SNPs have three possible genotypes

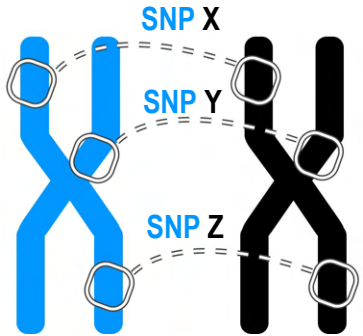
homozygous major, heterozygous, and homozygous minor

Epistasis detection and high-order searches

Looking into SNPs 1-by-1 does not uncover all genotype-phenotype associations

SNPs interact in non-linear ways
epistasis: trait is multi-SNP dependent

Computationally demanding
Especially if performing high-order searches



for each SNP combination

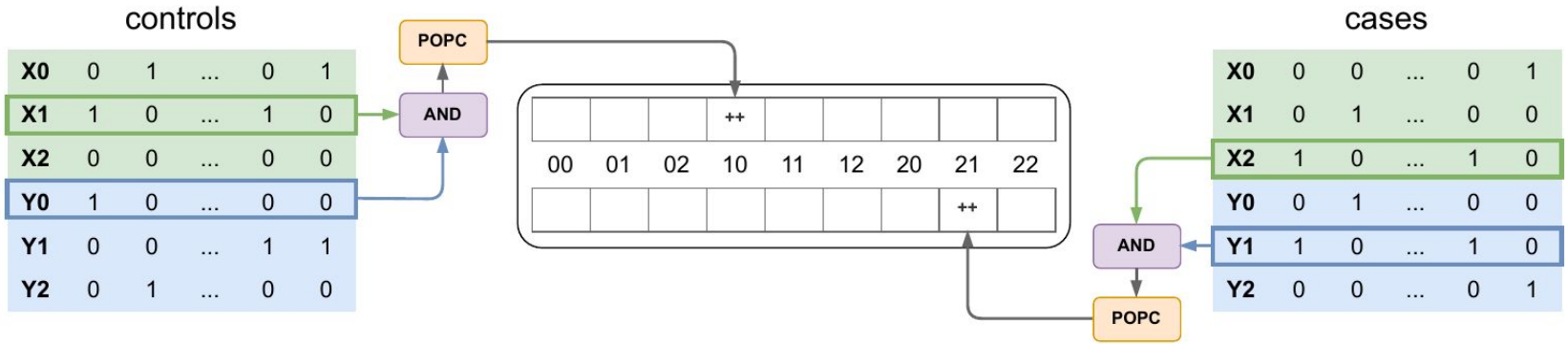
1. **Count genotypes**
2. **Apply scoring function**
3. **Reduce scores**

$$\frac{M!}{k!\,(M-k)!}$$

Challenge: Combinations grow exponentially with the number of SNPs (M) and interaction order (k)

Core operations for evaluating SNP combinations

Datasets often encode genotype using integers to denote the three possible alleles, i.e. homozygous major (0), heterozygous (1), and homozygous minor (2)



BOOST [1] introduced dataset binarization using 1-bit per SNP/genotype tuple

Genotype counts calculated applying k-1 AND per 1 POPC instruction to process sample packs

Use of GPUs and other parallel accelerators

GPUs are suitable for data-parallel apps
Makes GPUs a good choice for epistasis searches

Recent approaches tend to rely on GPUs
Although SoA uses also other devices (e.g. IPU)

Examples of GPU SoA approaches:
GBOOST [2]: extends BOOST to GPUs
MPI3SNP [3]: performs 3rd-order searches
CUDA-Episdet [4]: > 80% POPC peak
Cross-DPC-Episdet [5]: HW interoperability

GPUs also have native support for binary operations such as AND and POPC
POPC is slower than AND on GPUs, but use of bitwise operations still enables faster processing

Acceleration of core operations using tensor cores

CoMet [6] relies on 16-bit multiply-add on the tensor cores in Volta GPUs

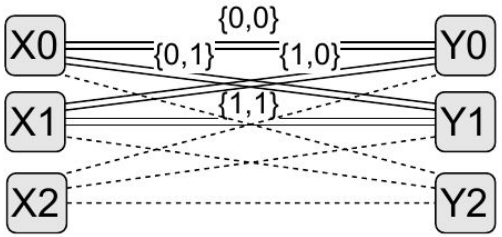
Processing blocks of data from multiple SNPs and samples using matrix multiplication

Tensor-Episdet [7] uses XOR+POPC on the tensor cores in Turing GPUs

Deriving AND+POPC at low cost and implementing technique that infers most genotypes

Genotype inference enables reconstructing full tables (2×3^k values) from 2×2^k genotype counts

2×4 in 2nd-order searches and 2×8 in 3rd-order searches



Tensor cores for third-order searches and beyond

Tensorized 1-bit processing enabled 18× higher 3rd-order performance per GPU

Despite the Titan RTX being similar in FP16 compute throughput to the Tesla V100 (≈130 TFLOPS)

Approach	(# nodes ×) GPU config.	Performance	Performance / node	Performance / GPU
CoMet [6]	(4373 ×) 6 Tesla V100	81611	18.66	3.11
Tensor-Episdet [7]	(1 ×) 1 Titan RTX	54.54	54.54	54.54

Performance: SNP combinations processed per second scaled to the sample size

Epi4Tensor [8] does 4th order using AND+POPC introduced in Ampere

Over 70% of the peak tensor throughput, but requirements impractical for full-scale searches

Porting GPU-accelerated codes to supercomputers

Processing potential of supercomputers is often achieved with GPU accelerators

Most recent TOP500 list ranks nine supercomputers with GPUs as the ten most powerful systems

Approach	Order(s)	Prog. model(s)	Targeted hardware
CUDA-Episdet [4]	2nd, 3rd	OpenMP, CUDA	Any NVIDIA GPU (Maxwell to Ampere)
Tensor-Episdet [7]	2nd, 3rd	CUDA	Turing GPUs (also supports Ampere)
Epi4Tensor [8]	4th	OpenMP, CUDA	Ampere GPUs (also supports Turing)
Crossarch-Episdet [9]	2nd, 3rd	OpenMP, SYCL	CPU/GPU with SYCL support via oneAPI

Some supercomputers (e.g. LUMI) use GPUs from vendors other than NVIDIA

Most codes have been developed in CUDA and target single-GPU/single-node systems

MeluXina GPU-accelerated partition
(200 ×) AMD EPYC 7452 + 4 × A100 SXM4 40GB

LUMI GPU-accelerated partition
(2978 ×) AMD EPYC 7A53 + 4 × AMD MI250X

NVIDIA A100 (MeluXina) and AMD MI250X (LUMI)

A100 SXM4 40GB

108 compute units @ 1.41 GHz
40 GB @ 1.56 TB/s

MI250X

2 × 110 compute units @ 1.70 GHz
2 × 64 GB @ 3.28 TB/s

The MI250X is a MCM with 2 GPU dies, seen from a software perspective as two GPUs

Both the A100 and the MI250X favour matrix in relation to vector throughput

The next-generation GPUs further amplify this gap, offering even greater speedups for matrix computations, which makes it increasingly important to prioritize matrix-based workloads

TFLOPS / TOPS

GPU / Precision	Vector FP64	Matrix FP64	Vector FP32	Matrix FP32	Vector FP16	Matrix FP16	Matrix INT8	Matrix INT4	Matrix INT1
A100	9.7	**19.5**	19.5	N/A	78	**312**	**624**	**1248**	**4992**
MI250X	47.9	**95.7**	47.9	**95.7**	N/A	**383**	**383**	**383**	N/A

Adding support for AMD GPUs through HIP

All three CUDA codes were converted to HIP for compatibility with LUMI

HIPIFY + microarchitecture specific modifications + adaptations to HIP/ROCm software stack

Tensor-Episdet and Epi4Tensor rely on CUTLASS for 1-bit on tensor cores

Scalability tests used FP16 via hipBLAS on the MI250X, which lacks tensorized 1-bit support

The MI250X supports integer ops at the same peak throughput as FP16

Precision tuning considering the INT8 data type, which is also supported through hipBLAS

Implementing intra/inter-parallelization with MPI

Both intra-node and inter-node parallelization are implemented via MPI

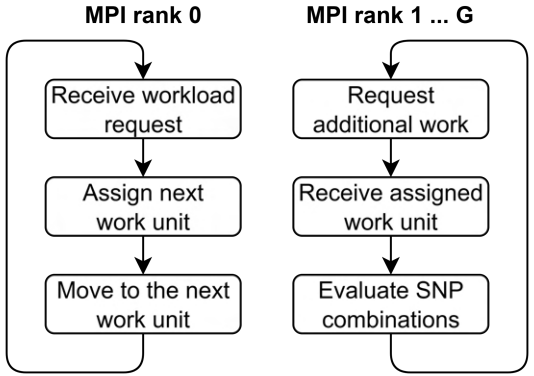
Performed at the level of loop iterations, with each iteration treated as a distinct work unit

Iterations in vector codes process the same number of combinations

Load balance achieved statically assigning work to as many MPI processes as there are GPUs

In matrix methods an SNP block pairs with itself and higher-index blocks

Dynamic scheduling via an extra MPI process handles the variable amount of combinations



Speedups and performance for single-node runs

Close to linear speedups, i.e. 4× on MeluXina and 8× on LUMI, are achieved

Using MPI to separately address each GPU die of the MI250X enabled its efficient utilization

Speedups for single-node runs with CUDA-Episdet (8192 SNPs), Crossarch-Epistasis (8192 SNPs), Tensor-Episdet (16384 SNPs) and Epi4Tensor (2048 SNPs). Tensor-Episdet and Epi4Tensor have been executed with 524288 samples on MeluXina, all other runs process 32768 samples.

Performance with tensor cores is higher than that of CUDA/stream core methods

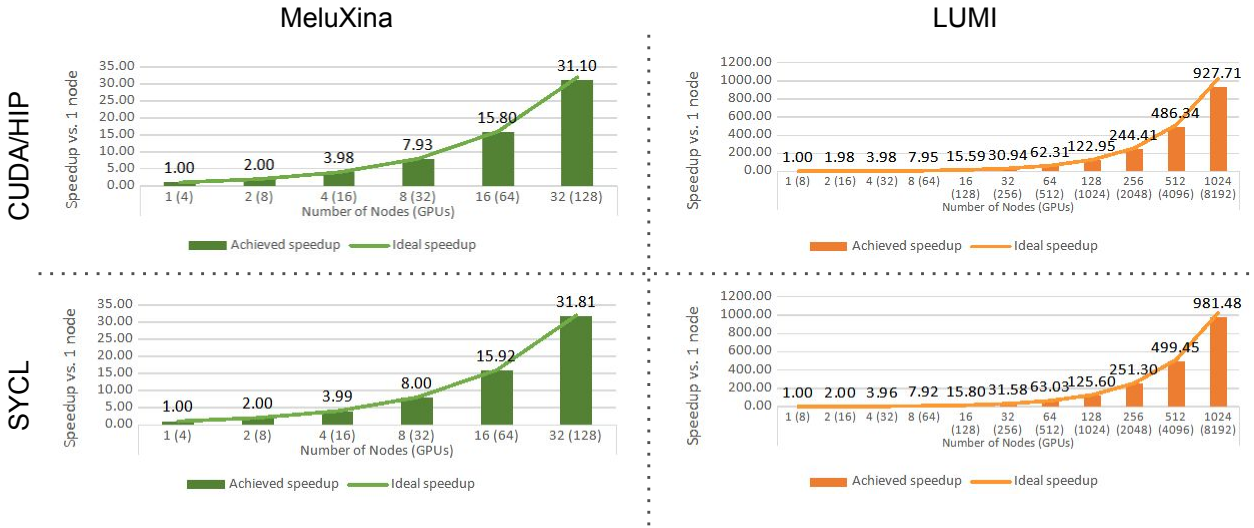
CUDA/HIP-Episdet: 10.96 (MeluXina) / 17.86 (LUMI); Crossarch-Episdet: 10.20 (MeluXina); 11.79 (LUMI)

Tensor-Episdet: 344.43 (MeluXina) / 21.06 (LUMI); Epi4Tensor: 429.58 (MeluXina) / 13.00 (LUMI)

Third-order searches using CUDA/stream cores

Runs on MeluXina and LUMI using up to 32 and 1024 GPU-accelerated nodes

Parallel efficiency of 96% on 1024 nodes for the SYCL code in comparison to single-node



Speedups with CUDA/HIP-Episdet and Crossarch-Episdet on multiple nodes (MeluXina: 16384 SNPs × 32768 samples, LUMI: 32768 SNPs × 32768 samples).

CUDA/HIP faster despite better scaling for SYCL

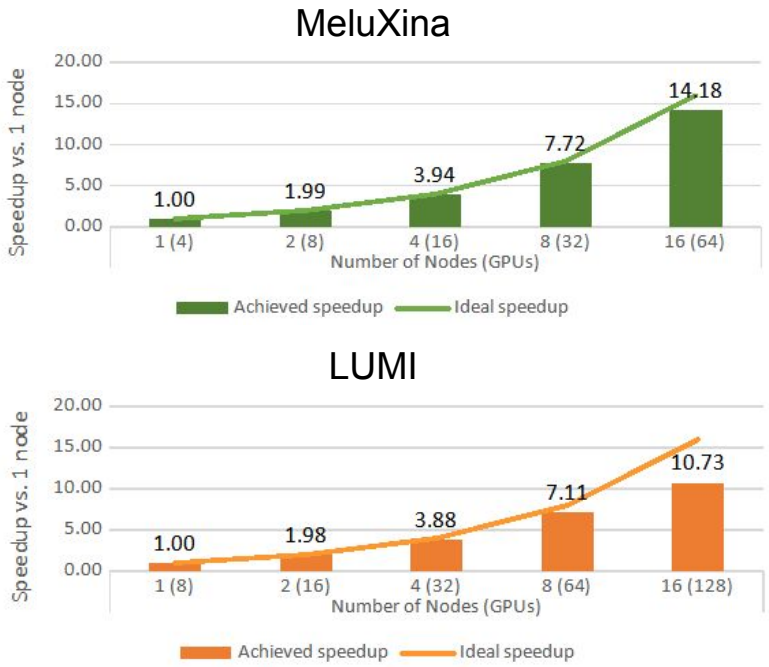
92.7% of CUDA w/ 32N

Perf: 317.56 vs. 342.61

70.1% of HIP w/ 1024N

Perf: 11182.24 vs. 15951.32

Third-order searches using matrix processing cores



Speedups with Tensor-Episdet on multiple nodes (MeluXina: 32768 SNPs × 524288 samples, LUMI: 32768 SNPs × 32768 samples).

Both LUMI and MeluXina scale similarly well with the number of targeted GPUs

LUMI does not scale so well per node because it has more GPU dies per node (8 instead of 4)

Significant performance gap due to the use of 1-bit (A100) vs. 16-bit (MI250X)

5212.88 (MeluXina) vs. 226.36 (LUMI) tera SNPs processed per second scaled to the sample size

Precision tuning allowed substantially increasing performance on AMD GPUs

1.6× faster using INT8 instead of FP16, despite the same theoretical throughput on the MI250X

Fourth-order searches using matrix processing cores

As in 3rd-order searches, 1-bit resulted in MeluXina achieving higher performance

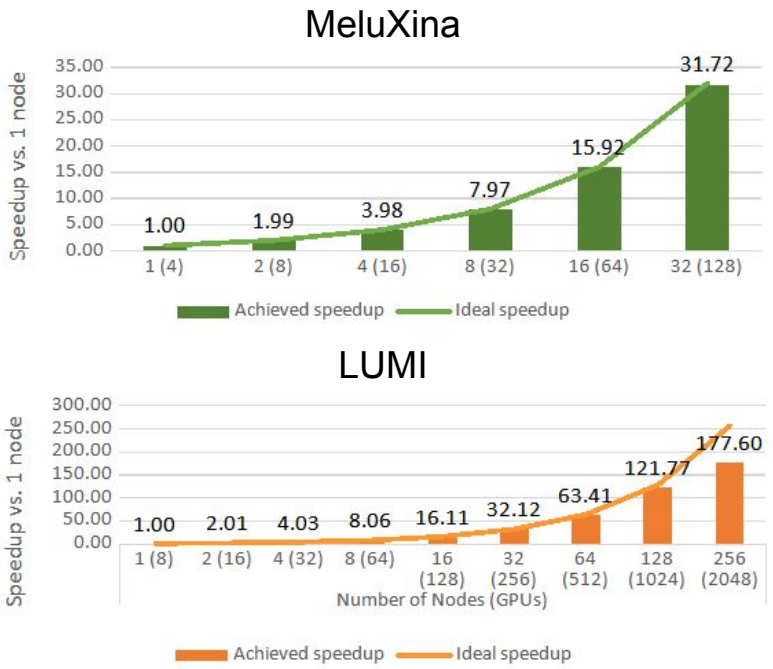
Up to 14346.33 (Meluxina) vs. 2456.96 (LUMI)

Runs with higher node counts on LUMI showcase the scalability of the method

Close to linear improvements up to 128 nodes

Scalability via 2D work decomposition

Avoids balancing work via smaller SNP blocks, which lowers tensor/matrix core throughput



Speedups with Epi4Tensor on multiple nodes (MeluXina: 4096 SNPs × 524288 samples, LUMI: 4096 SNPs × 32768 samples).

Conclusions and ongoing/future work

Ported four GPU-accelerated codes to two EuroHPC supercomputers

- All were parallelized to leverage multiple GPUs on multiple nodes
- Three went through CUDA-to-HIP translation targeting AMD GPUs

Highest-performing codes on AMD GPU accelerators and still improvable

- Additional tuning of SYCL code could reduce performance gap with HIP
- Matrix methods may benefit from further precision tuning, e.g. using 4-bit

MeluXina achieves higher performance on the codes using tensor cores

- Usage of more nodes on LUMI showcases scalability of the ported codes

Ongoing/Future work in the context of GPU-accelerated methods:

- Adapting our solutions to newer GPU microarchitectures (CDNA3, Hopper)
- Use of Joint Matrix SYCL extension as a means for higher interoperability

References

[1] X. Wan, A Fast Approach to Detecting Gene-Gene Interactions in Genome-wide Case-Control Studies, Am J Hum Genet., 2010

[2] L. Yung, GBOOST: a GPU-based tool for detecting gene-gene interactions in genome-wide case control studies, Bioinformatics, 2011

[3] C. Ponte-Fernández, Fast search of third-order epistatic interactions on CPU and GPU clusters, Int. J. High Perform. Comput. Appl., 2019

[4] R. Nobre, Accelerating 3-Way Epistasis Detection with CPU+GPU Processing, JSSPP 2020

[5] D. Marques, Unlocking Personalized Healthcare on Modern CPUs/GPUs: Three-way Gene Interaction Study, IPDPS 2022

[6] W. Joubert, Attacking the Opioid Epidemic: Determining the Epistatic and Pleiotropic Genetic Architectures for Chronic Pain and Opioid Addiction, SC18

[7] R. Nobre, Retargeting Tensor Accelerators for Epistasis Detection, TPDS, 2021

[8] R. Nobre, Tensor-Accelerated Fourth-Order Epistasis Detection on GPUs, ICPP22

[9] R. Nobre, Cross-architecture high-order exhaustive epistasis detection on CPU and GPU devices, Intel DevMesh, 2020


inesc id
lisboa
Thank you!

Unsupervised AI models for Down syndrome neuroimaging research: from automatic brain alteration detection to the generation of synthetic 3D MRI

Xavier Sevillano
Human-Environment Research Group
La Salle – Universitat Ramon Llull, Barcelona

EuroHPC Joint Undertaking

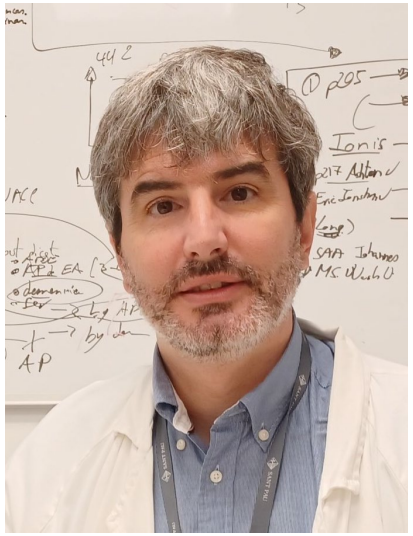
Project ID EHPC-AI-2024A02-043
EuroHPC User Days 2025
30 Sept - 1 Oct, Copenhagen

laSalle
UNIVERSITAT RAMON LLULL

A multidisciplinary research

Dr Neus Martínez-Abadías

UNIVERSITAT DE BARCELONA



Dr Juan Fortea

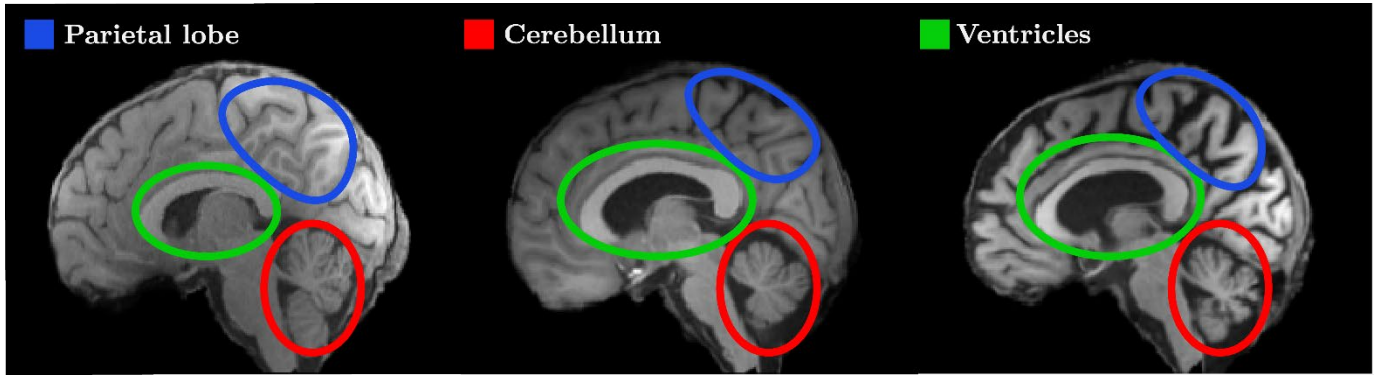
Sant Pau memory unit

Outline

- Down syndrome
- Brain magnetic resonance imaging
- AI applied to neuroimaging
- Towards brain biomarkers discovery
- Representation learning for patient stratification
- Generation of synthetic 3D brain MRI scans
- Conclusions
- Scientific outputs
- Acknowledgments

Down syndrome

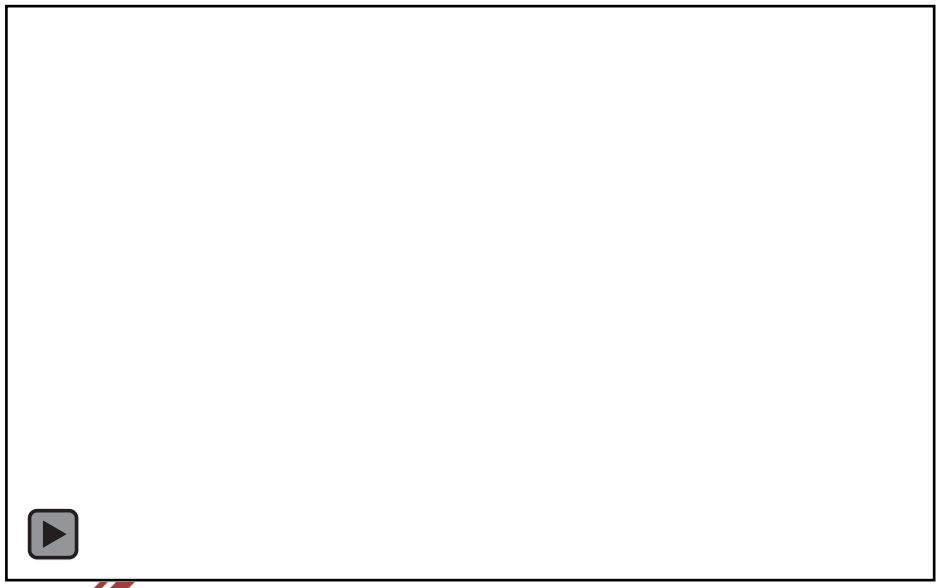
- Down syndrome (DS) is a **complex genetic disorder** caused by the triplication of chromosome 21, occurring in 1 out of every 700 to 1,000 live births
- Individuals with DS present a characteristic **craniofacial phenotype that constrains the growth and shape of the brain**, with disproportionately smaller hippocampus and cerebellum, large ventricles and other alterations

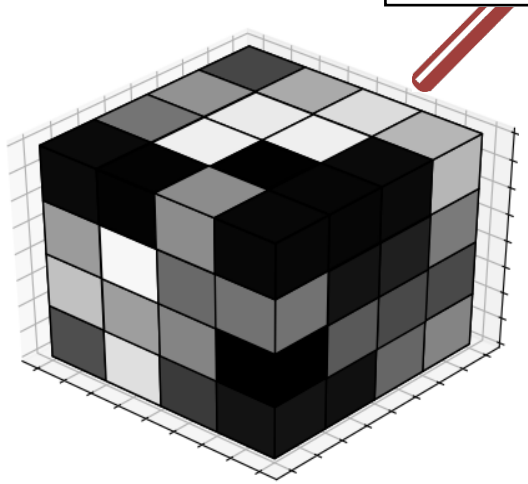


- The structural brain alterations in DS have been associated with cognitive and functional disabilities from birth, and early dementia in adulthood: **after age 40, people with DS are at ultra-high risk of developing Alzheimer's disease** (AD) due to accelerated neurodegeneration

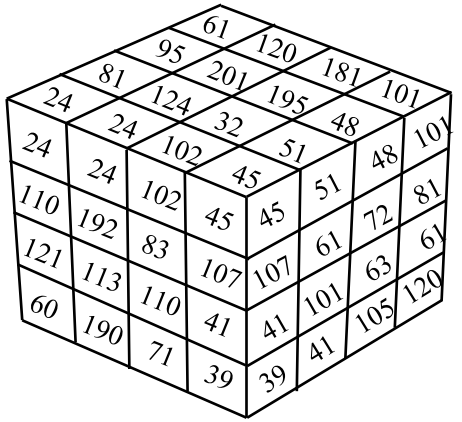
Brain magnetic resonance imaging

- Among neuroimaging techniques, **brain magnetic resonance imaging (MRI) is the gold standard** for its non-invasive, radiation-free nature and superior soft tissue contrast

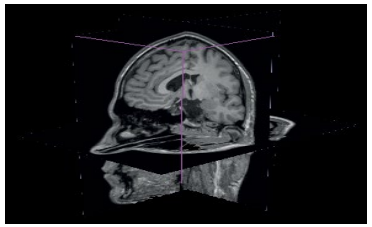




3D voxel-based representation



AI applied to neuroimaging



Deep Learning Model

Brain registration

Brain segmentation

Diagnostic classification

Training

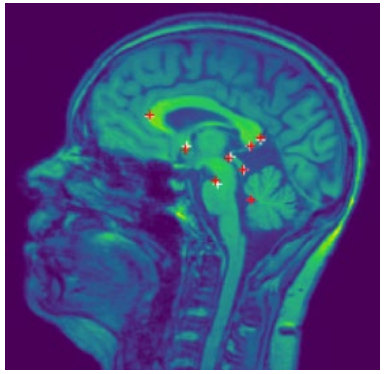
But… where can we get all this data from?

"Annotated" data

ADNI Alzheimer's Disease Neuroimaging Initiative

ABC-DS Alzheimer Biomarker Consortium-Down Syndrome

uk biobank Enabling scientific discoveries that improve human health



"Single-Landmark vs. Multi-Landmark Deep Learning Approaches to Brain MRI Landmarking: a Case Study with Healthy Controls and Down Syndrome Individuals"
(British Machine Vision Conference 2023)

AI applied to neuroimaging

- Due to their unsupervised nature, generative models avoid the need for large, annotated datasets of brain MRI scans

- Scientific questions:
  - Can generative models …
    - contribute to the discovery of brain biomarkers of DS and AD in DS?
    - stratify DS individuals depending on their degree of neurodegeneration and other comorbidities?
    - allow the generation of realistic synthetic 3D brain imaging?

- Unsupervised models:
  - Autoencoders
  - Diffusion models

MONAI PyTorch

- Deployed on Marenostrum 5 ACC nodes at the Barcelona Supercomputing Center

BSC

AI applied to neuroimaging

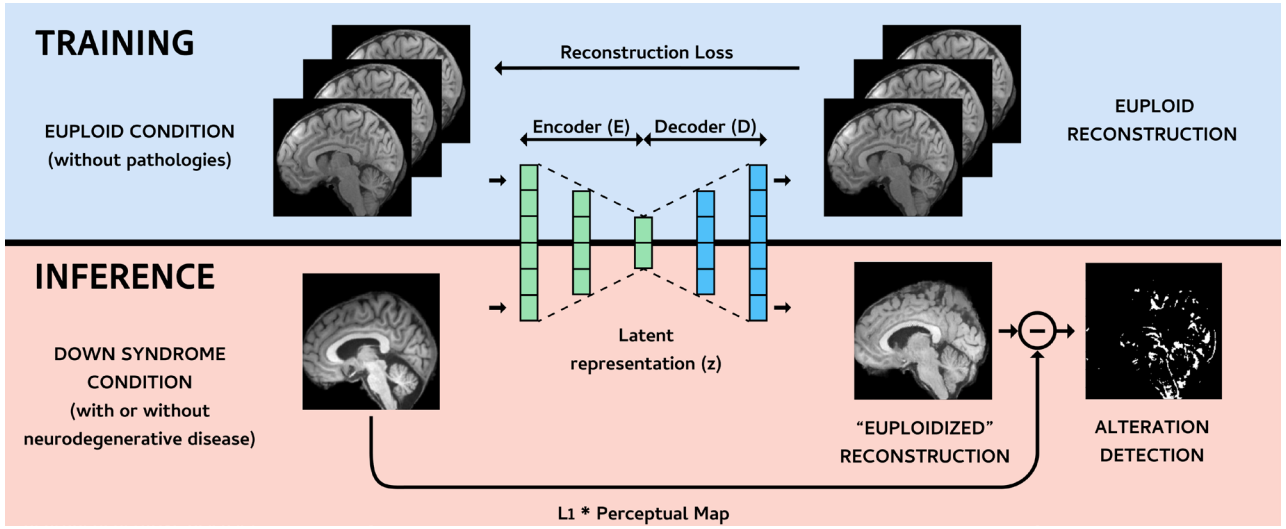
- Data used in the project

Dataset	IXI	HCP	Sant Pau	ABC-DS
Usage	VAE Training	VAE Training	Classification	Classification
Nº of subjects	580	1113	931	63
Diagnosis	EU - HC	EU - HC	EU(540) / DS(391)	DS

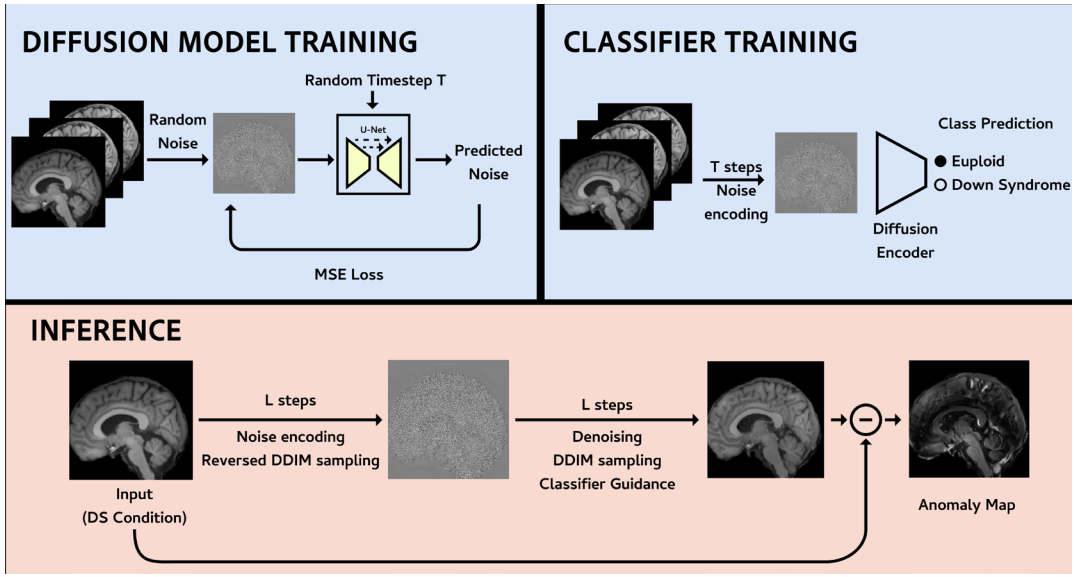
Dataset	Sant Pau Down syndrome subjects					
Usage	Intellectual disability level in DS			Alzheimer's disease progression in DS		
Diagnosis	Mild	Moderate	Severe	No deterioration	Prodromic	Established
Nº of subjects	114	214	59	218	36	108

Towards brain biomarkers discovery

- With autoencoders

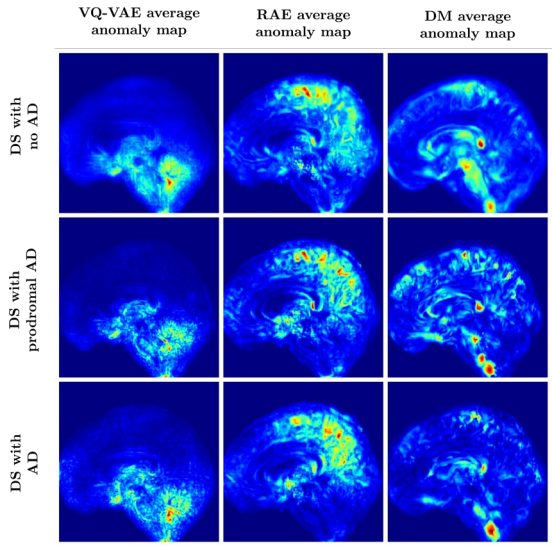


- With diffusion models



Towards brain biomarkers discovery

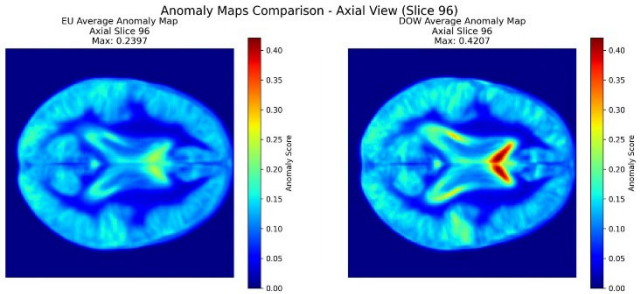
- Before access to HPC infrastructure (Tesla V100 GPUs): limited to 2D or 2.5D

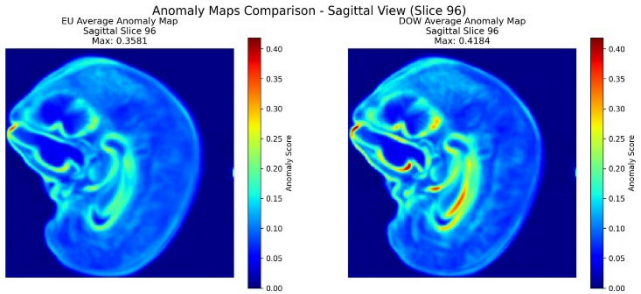


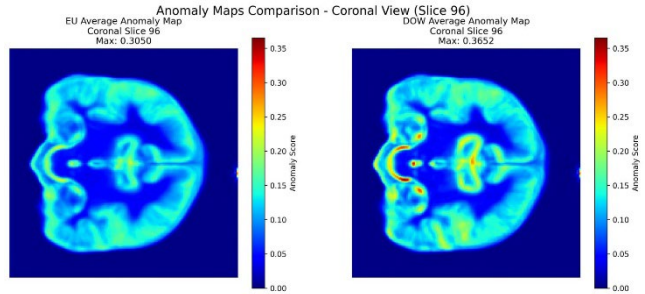
"Towards the Discovery of Down Syndrome Brain Biomarkers Using Generative Models"
(BICW, European Computer Vision Conference 2024)

Towards brain biomarkers discovery

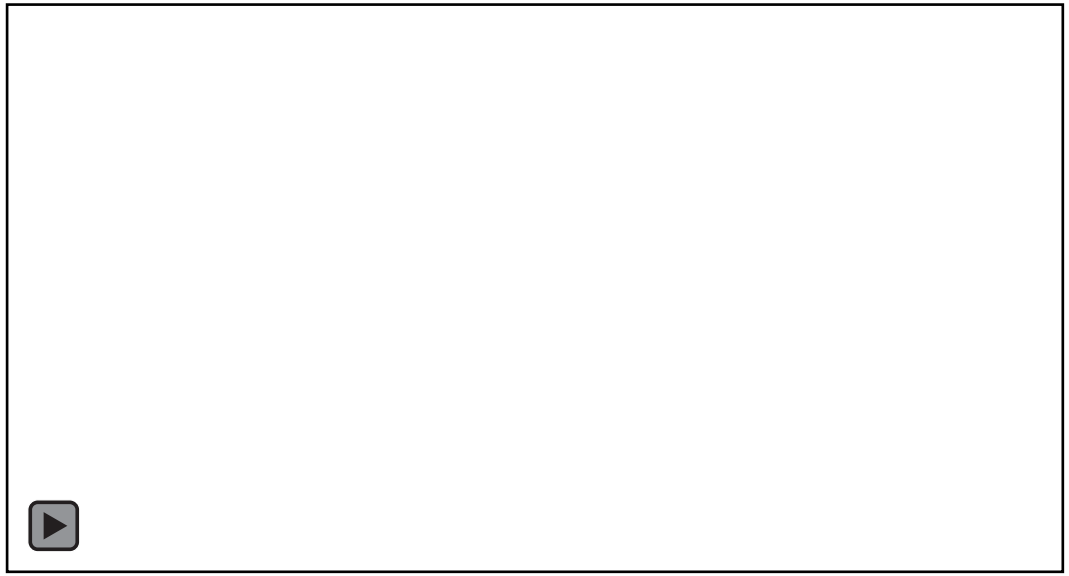
- Variational Autoencoders to compute 3D brain alteration maps between euploid (EU) and DS populations
  - Three sectional views of EU vs EU and DS vs EU anomaly maps





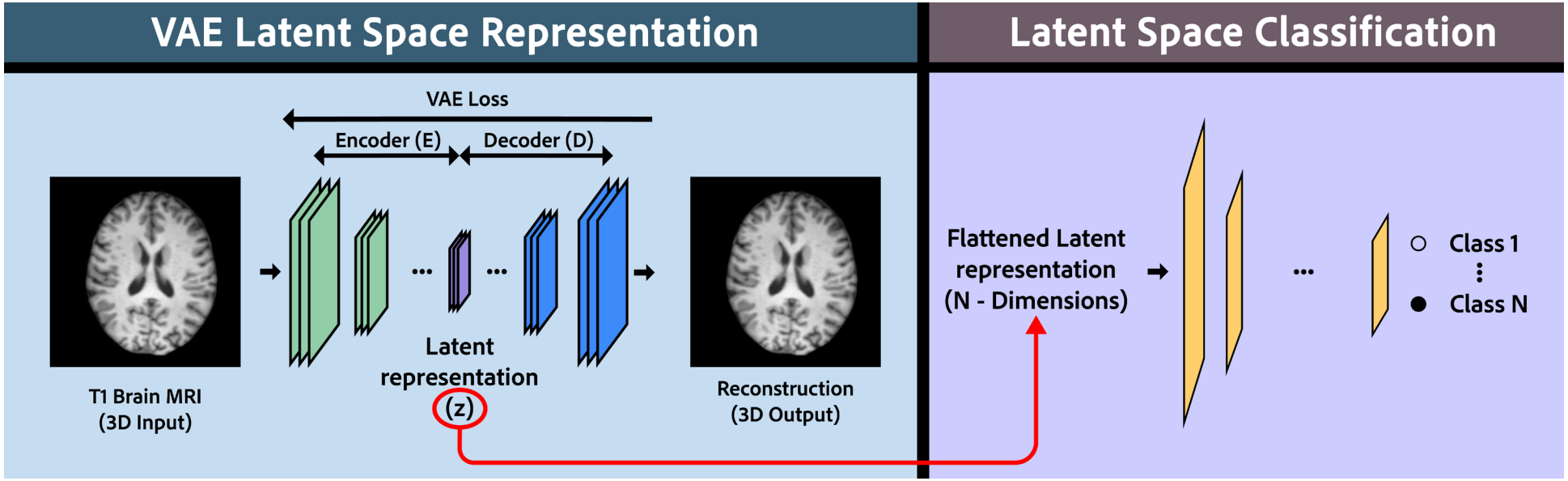


  - Full 3D average brain alteration map between DS and EU populations



Representation learning for patient stratification

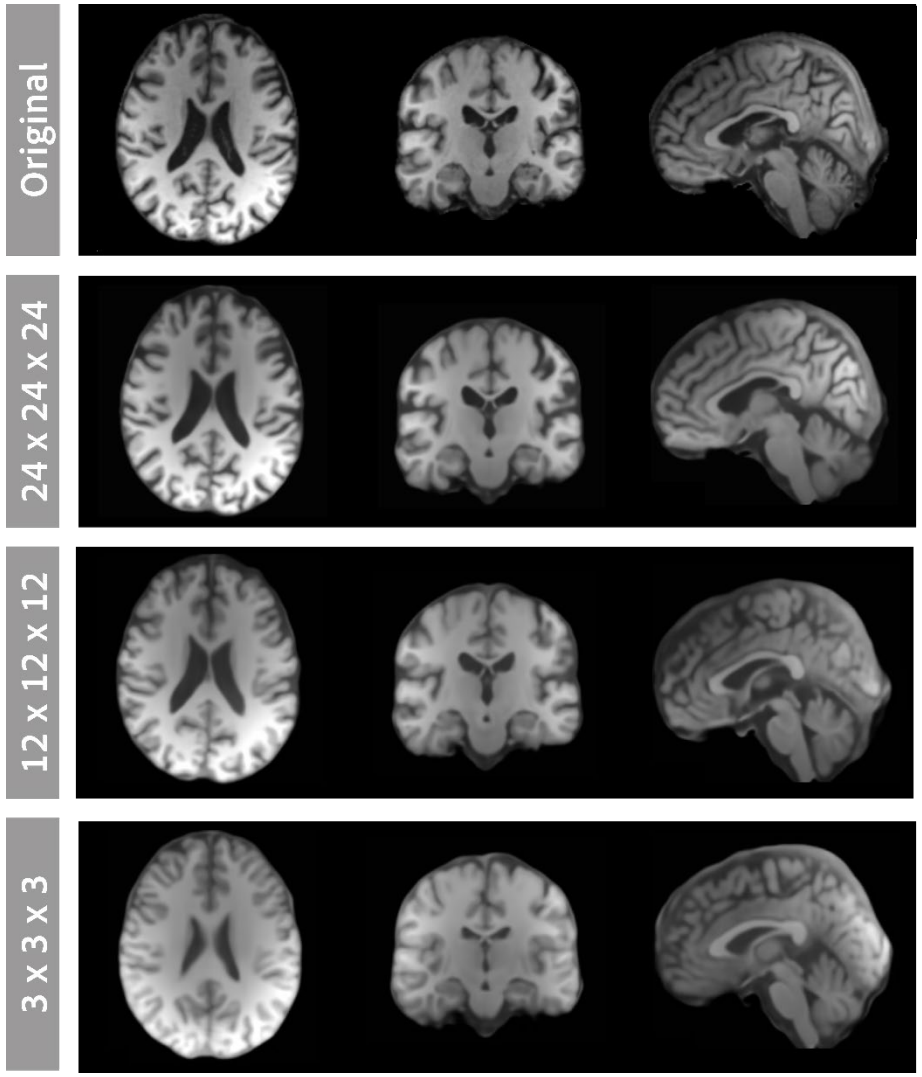
- Study if latent representations learnt by 3D autoencoders are features with diagnostic potential



- Evaluation:
  - Qualitative: reconstruction fidelity and embedding interpretability
  - Quantitative: performance on different classification tasks

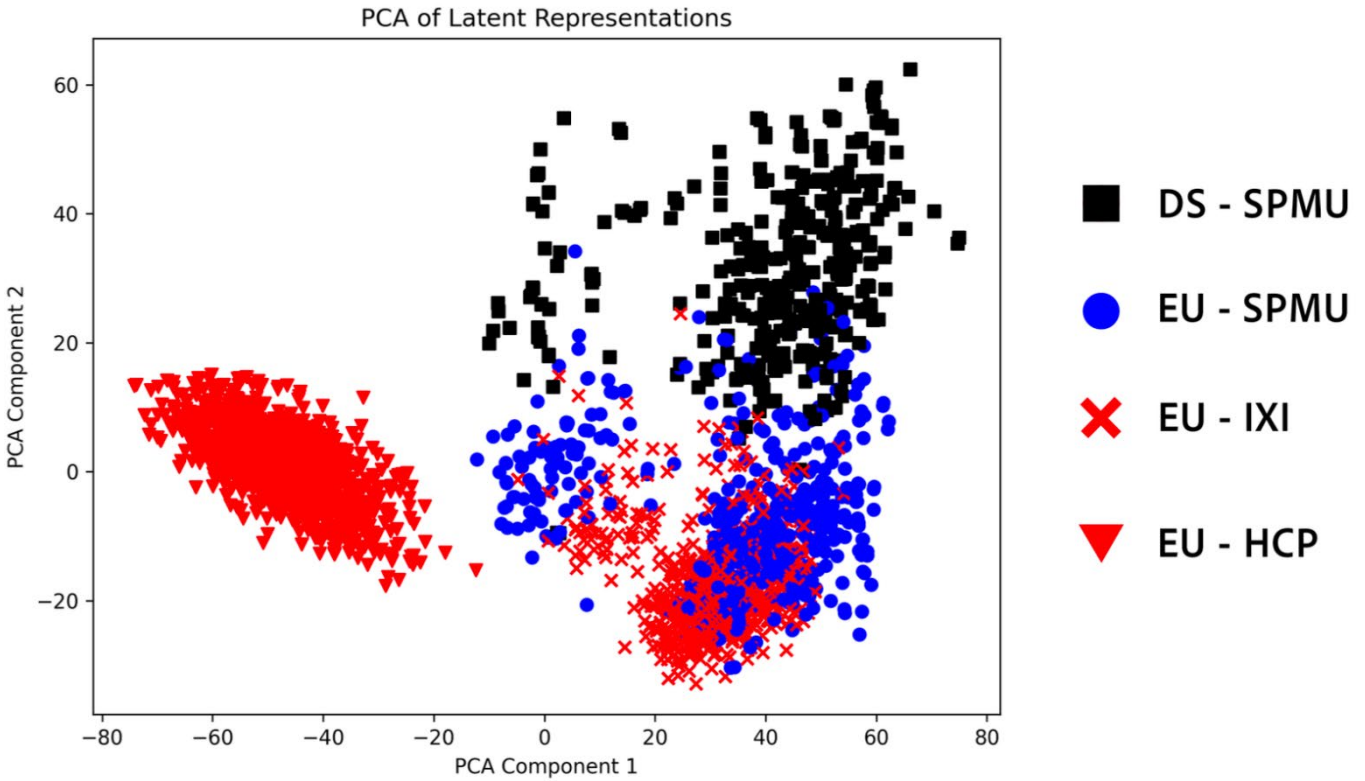
Representation learning for patient stratification

- Reconstruction fidelity in terms of latent space dimensionality



Representation learning for patient stratification

- Embedding interpretability



Representation learning for patient stratification

- Classification performance
  - Task 1: Euploid vs Down syndrome

Sant Pau dataset

Latent Space	PCA	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
(24, 24, 24)	-	99.4	99.8	98.7	0.99
(24, 24, 24)	2	95.3	95.7	94.9	0.98

Generalisation on ABC-DS dataset

Latent Space	PCA	Accuracy (%)
(24, 24, 24)	-	99.7

Representation learning for patient stratification

- Classification performance

  - Task 2: Intellectual disability stratification

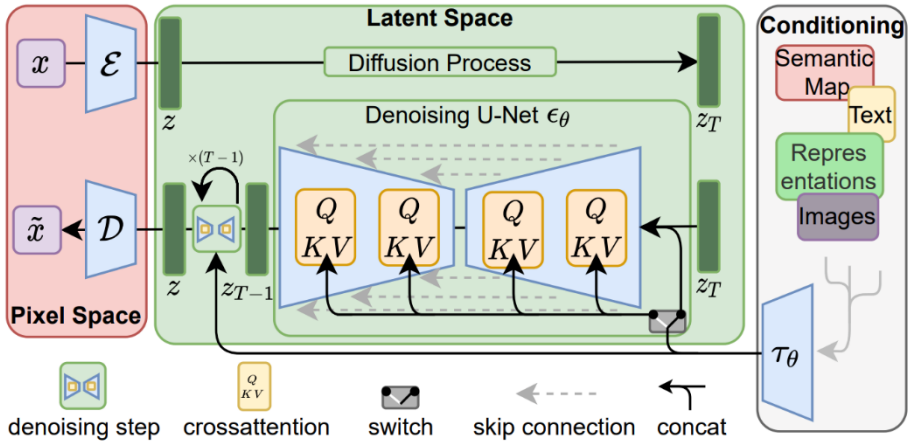
Accuracy (%)	Sensitivity (%)			Specificity (%)			AUC
	Mild	Moderate	Severe	Mild	Moderate	Severe	
70.0 ± 3.0	44.0	96.8	27.4	95.6	39.9	100	0.89

  - Task 3: Alzheimer's disease progression

Accuracy (%)	Sensitivity (%)			Specificity (%)			AUC
	No deter.	Prodromic	Established	No deter.	Prodromic	Established	
76.0 ± 5.0	94.0	11.1	61.0	61.7	96.9	91.3	0.82

Generation of synthetic 3D brain MRI scans

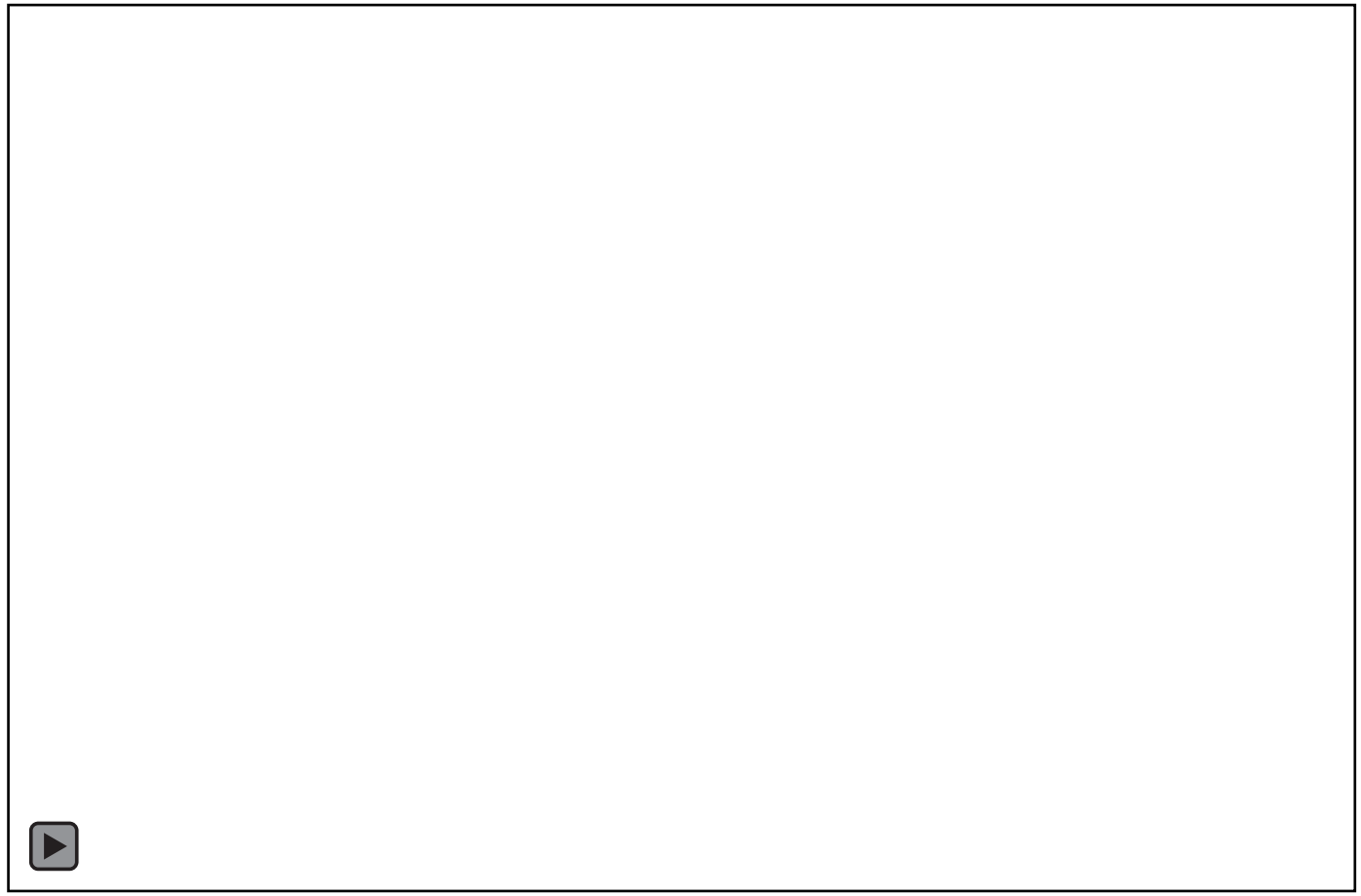
- Diffusion Models (DM) can generate high-quality synthetic image data (Midjourney, DALL-E, Imagen…)
- The high dimensionality of 3D medical imaging makes diffusion prohibitive on voxel space $\rightarrow$ Latent Diffusion Models (LDM)
- Latent representations are learned via pretrained autoencoders



Rombach et al., 2022 - *High-Resolution Image Synthesis with Latent Diffusion Models*

- LDMs can be conditionally guided, for constrained data generation

Generation of synthetic 3D brain MRI scans



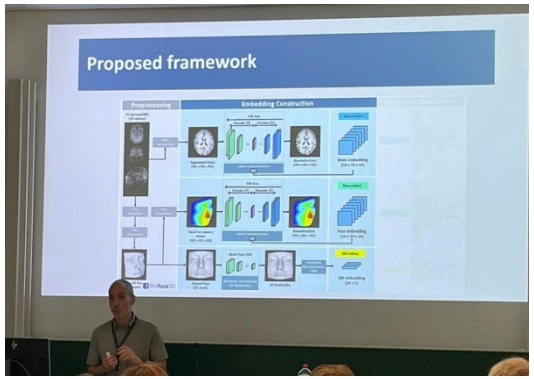
Synthetic 3D brain MRI scan generated by LDM

Conclusions

- Access to HPC has made whole-3D brain MRI processing possible
- No experience in HPC: overestimation of computational resources
- With the support of EPICURE team, we could apply Data Distributed Parallelism strategies on the autoencoders training, scaling the number of nodes from 1 to 8
- Strongly limited by the availability of data
- Deeper integration with clinical partners

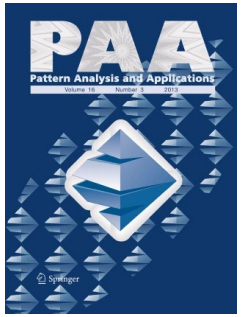
Scientific outputs

- Congress participations
  - 2nd ADAD-DSAD Conference (Barcelona, Spain)
  - 3rd Facial Genetics Symposium (Leuven, Belgium)



- Peer-reviewed conference paper
  - 12th Iberian Conference on Pattern Recognition and Image Analysis (Coimbra, Portugal)

- Journal article
  - Pattern Analysis and Applications



Acknowledgments

Jordi Malé

Andrea Labá

laSalle
UNIVERSITAT RAMON LLULL

- EPICURE team at BSC (Guillem Cortiada, Alexandros Paliouras, David Vicente and Gaurav Saxena)
- EuroHPC JU for awarding the project ID EHPC-AI-2024A02-043 access to MareNostrum5 ACC at BSC

EuroHPC
Joint Undertaking

BSC

Unsupervised AI models for Down syndrome neuroimaging research: from automatic brain alteration detection to the generation of synthetic 3D MRI

Xavier Sevillano
Human-Environment Research Group
La Salle – Universitat Ramon Llull, Barcelona

EuroHPC Joint Undertaking

Project ID EHPC-AI-2024A02-043
EuroHPC User Days 2025
30 Sept - 1 Oct, Copenhagen

laSalle
UNIVERSITAT RAMON LLULL