Outline

- sbv IMPROVER concept

- Computational challenges open to the community

- The science in the challenge

- The sbv IMPROVER Israel Epigenomics challenge

PMI RESEARCH & DEVELOPMENT

Outline

- sbv IMPROVER concept

- Computational challenges open to the community

- The science in the challenge

- The sbv IMPROVER Israel Epigenomics challenge

PMI RESEARCH & DEVELOPMENT

IMPROVER: Industrial Methodology for Process Verification in Research

Project initiated 6 years ago and funded by Philip Morris International

Aims to provide a measure of quality control in R&D by identifying the building blocks that need verification in a complex industrial research pipeline

Aims to verify methods & data in systems biology / toxicology using double blind performance assessment

Complements the classical peer review system

_computational BIOLOGY

COMMENTARY

Verification of systems biology research in the age of collaborative competition

Pablo Meyer[1], Leonidas G Alexopoulos[2], Thomas Bonk[3], Andrea Califano[4], Carolyn R Cho[5], Alberto de la Fuente[6], David de Graaf[7], Alexander J Hartemink[8], Julia Hoeng[3], Nikolai V Ivanov[3], Heinz Koeppl[9], Rune Linding[10], Daniel Marbach[11], Raquel Norel[1], Manuel C Peitsch[3], J Jeremy Rice[1], Ajay Royyuru[1], Frank Schacherer[12], Joerg Sprengel[13], Katrin Stolle[3], Dennis Vitkup[4] & Gustavo Stolovitzky[1]

Collaborative competitions in which communities of researchers compete to solve challenges may facilitate more rigorous scrutiny of scientific results.

Nature Biotechnology 2011 Sep 8;29(9):811-5

BIOINFORMATICS   REVIEW   Vol. 28 no. 9 2012, pages 1193–1201 doi:10.1093/bioinformatics/bts116

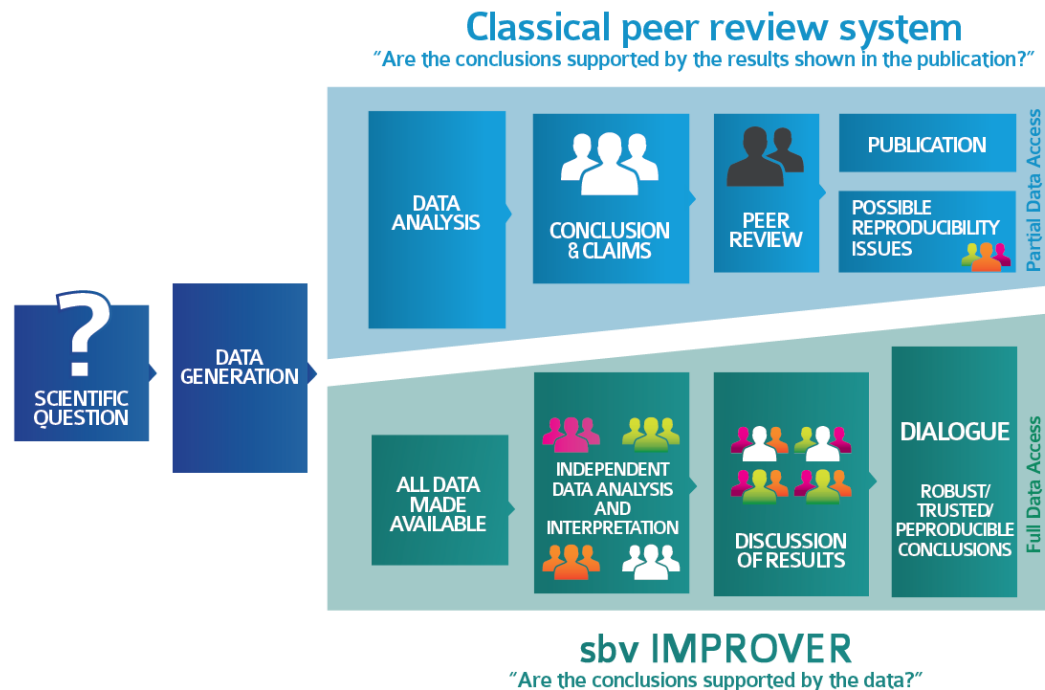Systems biology   Advance Access publication March 14, 2012

Industrial methodology for process verification in research (IMPROVER): toward systems biology verification

Pablo Meyer[1,†], Julia Hoeng[2,†], J. Jeremy Rice[1,†] Raquel Norel[1], Jörg Sprengel[3], Katrin Stolle[2], Thomas Bonk[2], Stephanie Corthesy[3], Ajay Royyuru[1,*], Manuel C. Peitsch[2,*] and Gustavo Stolovitzky[1,*]

[1]IBM Computational Biology Center, Yorktown Heights, 10598 NY, USA, [2]Phillip Morris Products SA, Research and Development, 2000, Neuchâtel, Switzerland and [3]IBM Life Sciences Division,8802, Zurich, Switzerland
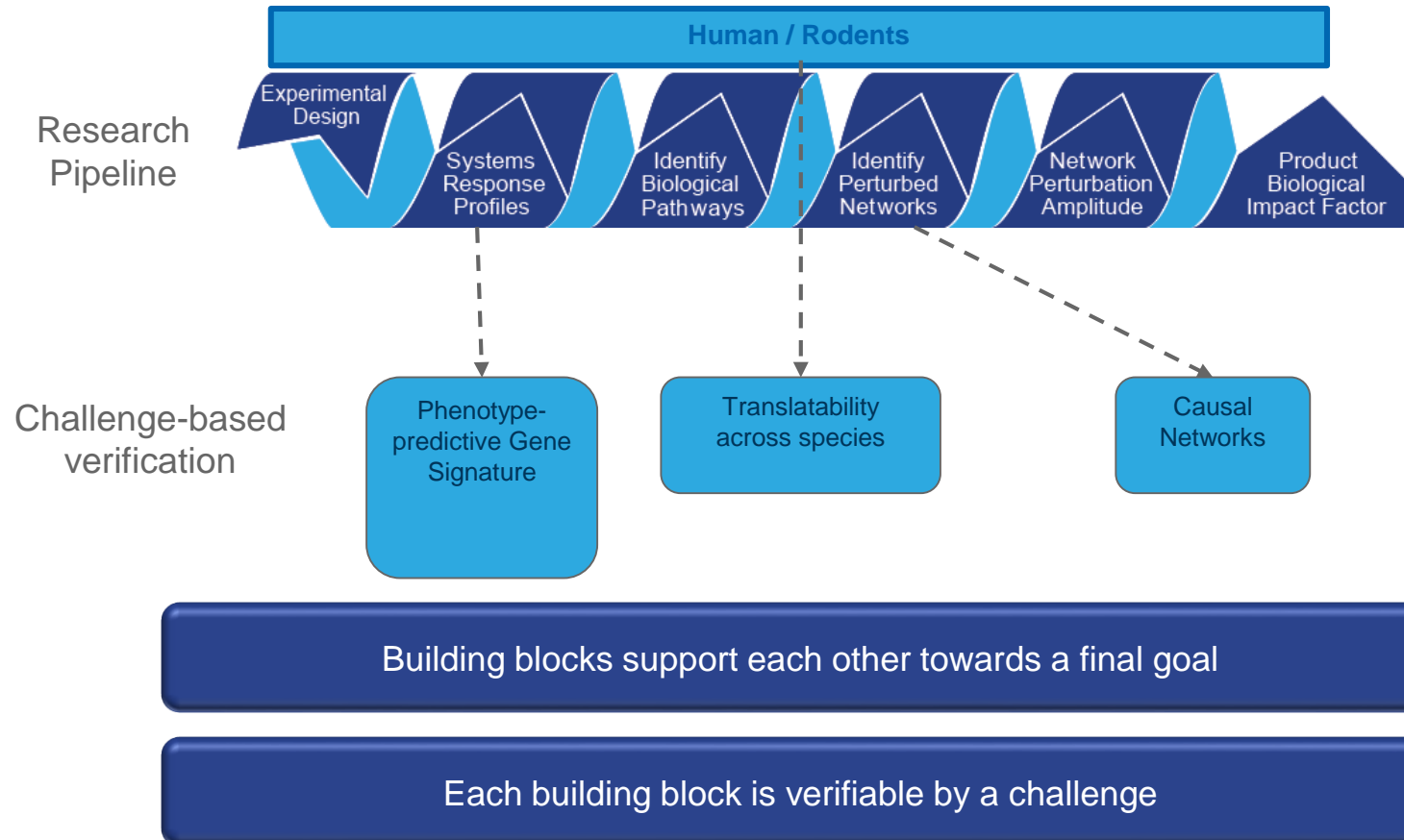
Bioinformatics 2012 28(9):1193-1201

PMI RESEARCH & DEVELOPMENT

sbv IMPROVER leverages the crowd to complement the classical peer review system

Classical peer review system
"Are the conclusions supported by the results shown in the publication?"

SCIENTIFIC QUESTION → DATA GENERATION → DATA ANALYSIS → CONCLUSION & CLAIMS → PEER REVIEW → PUBLICATION / POSSIBLE REPRODUCIBILITY ISSUES
Partial Data Access

ALL DATA MADE AVAILABLE → INDEPENDENT DATA ANALYSIS AND INTERPRETATION → DISCUSSION OF RESULTS → DIALOGUE / ROBUST/ TRUSTED/ REPRODUCIBLE CONCLUSIONS
Full Data Access

sbv IMPROVER
"Are the conclusions supported by the data?"

- **Crowd-sourcing**: A natural evolution of web technologies led to the development of distributed problem-solving. Challenges are broadcasted to potential interested stakeholders (solvers). The winning participants are rewarded either with monetary awards, prizes, certificates, or with recognition.

- **Collaboration by Competition**: The scientific community sought to understand the limitations and comparative advantages of their methods by challenging model developers to make blind predictions on previously unseen data in a competitive framework.

- The community appreciates the successful methods which grow in credibility. Therefore, consideration of the scientific community is one of the forces that shape what is **currently considered as the way to do the science right**

PMI RESEARCH & DEVELOPMENT

Complex industrial research pipeline/workflow divided into verifiable building blocks

Human / Rodents

Research Pipeline

Experimental Design

Systems Response Profiles

Identify Biological Pathways

Identify Perturbed Networks

Network Perturbation Amplitude

Product Biological Impact Factor

Challenge-based verification

Phenotype-predictive Gene Signature

Translatability across species

Causal Networks

Building blocks support each other towards a final goal

Each building block is verifiable by a challenge

PMI RESEARCH & DEVELOPMENT

Outline

- sbv IMPROVER concept

- **Computational challenges open to the community**

- The science in the challenge

- The sbv IMPROVER Israel Epigenomics challenge

PMI RESEARCH & DEVELOPMENT

How to build a challenge

PMI RESEARCH & DEVELOPMENT
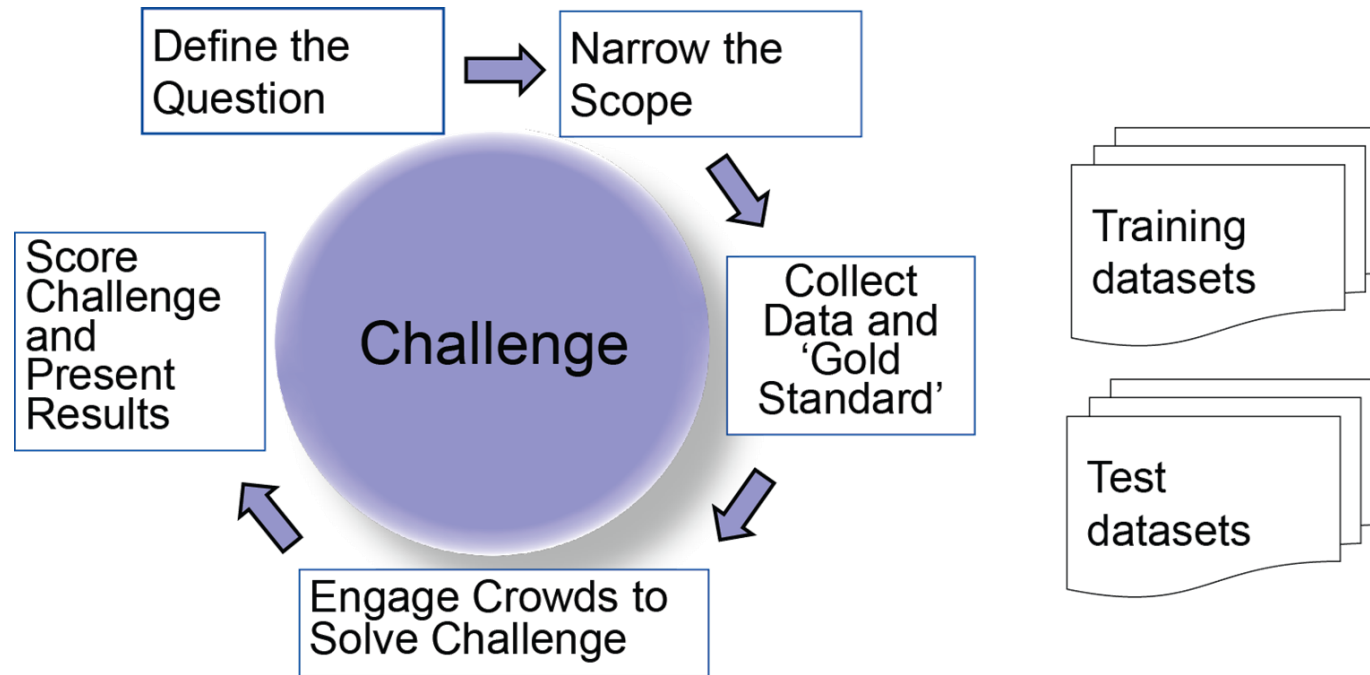
Past sbv IMPROVER computational challenges

Diagnostic signature challenge (2012):
- Verify computational approaches that classify clinical samples based on transcriptomics data in 4 disease areas

Species translation challenge (2013):
- Identify a function which maps measurements derived from systematic perturbations in one species to another
- Understand the system boundaries of the translatability concept
- Quantify the translatability between species
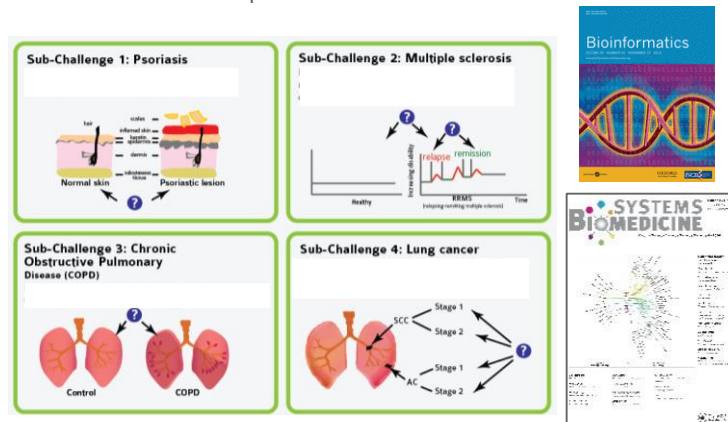
Network verification challenge (2014-2015)
- Engage the scientific community in the review of biological network models that are suitable for drug discovery, toxicological and mechanistic research in respiratory disease

Systems Toxicology Computational challenge (2015-2016)
- Verify that a robust predictive signature can be extracted from gene expression data that differentiates smokers, former smokers, and never smoker subjects
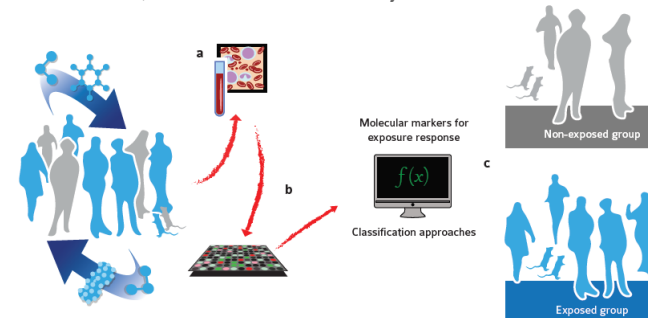
PMI RESEARCH & DEVELOPMENT

Outline

- sbv IMPROVER concept

- Computational challenges open to the community

- **The science in the challenge**

- The sbv IMPROVER Israel Epigenomics challenge

PMI RESEARCH & DEVELOPMENT

DNA methylation and smoke

- **Heavy smokers** can have **DNA methylation** changes in the upper respiratory tract (Lee et al., 2013; Zöchbauer-Müller et al., 2003; Zöchbauer-Müller et al., 2008; Bhutani et al., 2008)

- Two CpG sites located on the gene body of the **aryl hydrocarbon receptor repressor gene** (AHRR) were reported to be **hypomethylated** in several biological samples from smokers (Dogan et al., 2014; Monick et al., 2012; Philibert et al., 2012-2013)

- DNA methylation on AHRR gene is likely to be reversible after **cessation** and it has even been proposed as a quantitative **biomarker** of smoking cessation (Zeilinger et al., 2013; Philibert et al., 2016)

PMI RESEARCH & DEVELOPMENT

DNA methylation and smoke

- **All the previous studies reported a weak effect of cigarette smoking on DNA methylation**

Breitling et al, 2011
Human peripheral blood cells
177 samples
27K BeadChips Illumina
Hypomethylation of **1 single CpG**

Joehanes et al, 2016
Infinium HumanMethylation 450 BeadChip
Blood / CD4 T cells
15 907 blood-derived DNA samples
16 cohorts (2433 current, 6518 former, and 6956 never smokers)
2623 CpG / 1405 genes

Shenker et al, 2012
Blood
374 samples
450K BeadsChIP illumina
Hypomethylation **9 CpGs**

Review:Talikka et al, 2012
Genomic impact of cigarette smoke, with application to three smoking-related diseases.
Crit Rev Toxicol. 2012 Nov;42(10):877-89

PMI RESEARCH & DEVELOPMENT

DNA methylation : Three distinct classes
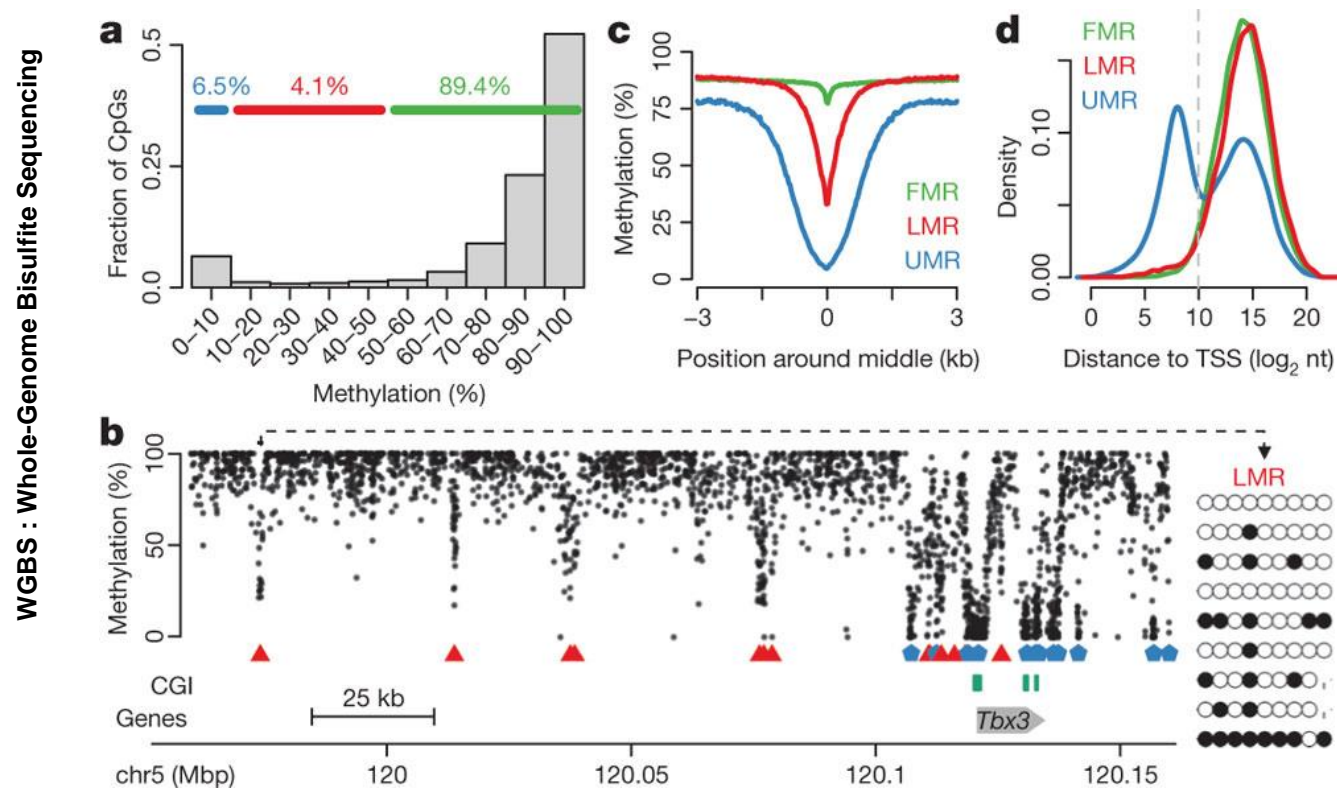
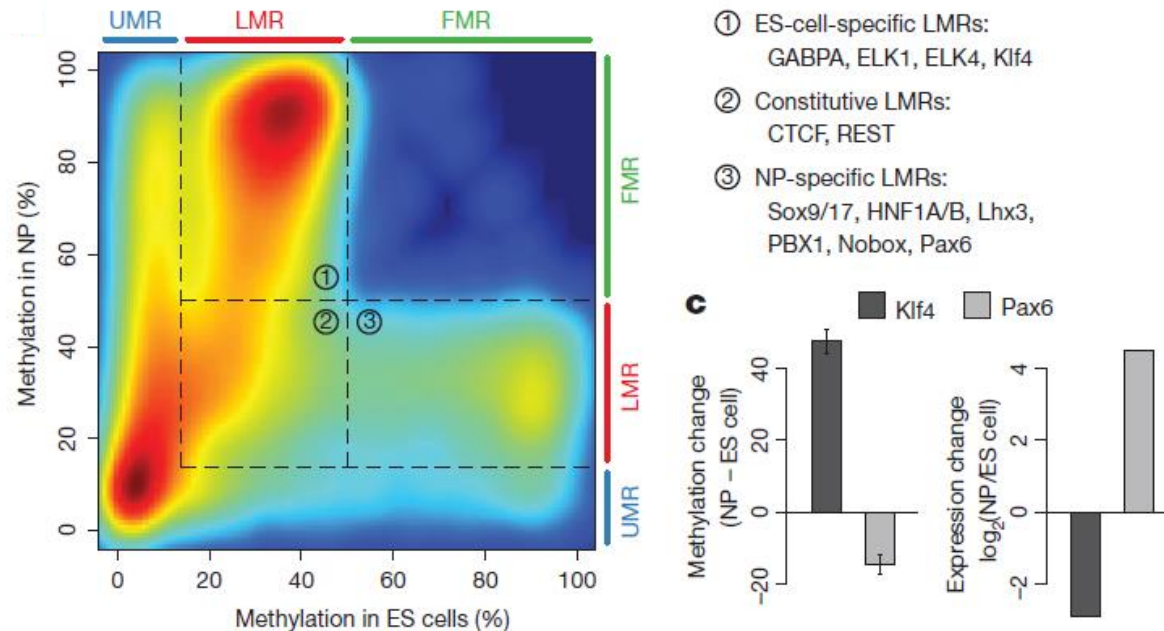UMRs: UnMethylated Regions   LMRs: Low Methylated Regions   FMR : Fully Methylated Regions

Stadler et al. Nature 2011 DNA-binding factors shape the mouse methylome at distal regulatory regions

PMI RESEARCH & DEVELOPMENT

DNA methylation provides useful information about the genomic context

- Low methylation level or no methylation = TF binding = Transcriptional activity

- High methylation level = Absence of TF binding = Absence of transcriptional activity, except for gene bodies
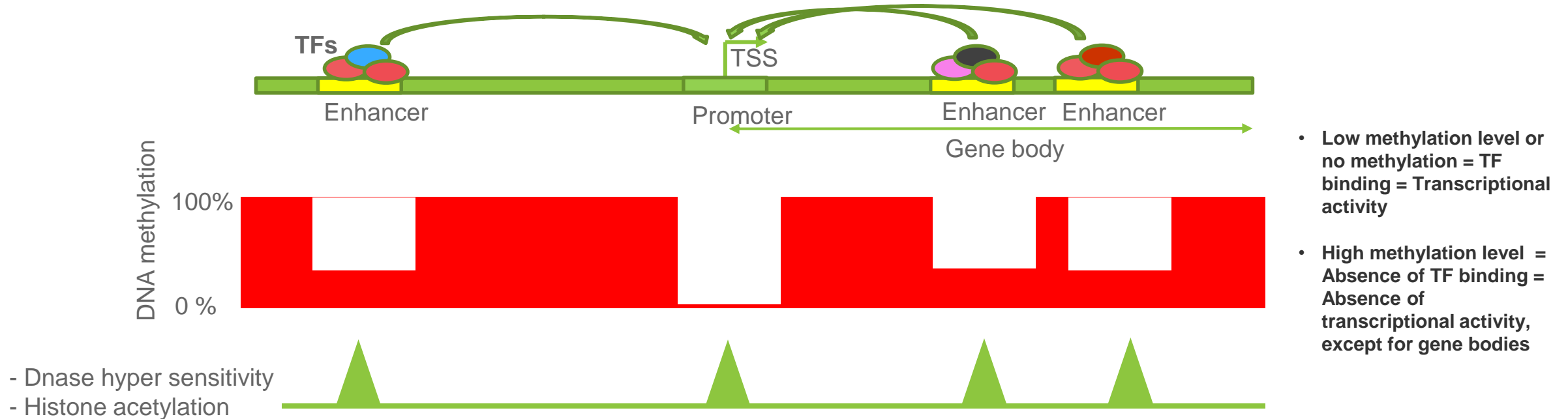
- The majority of CpGs in mammalian genomes are fully methylated. Hypomethylated CpGs reside at active Cis Regulatory elements (**CRE**s). Active **promoters** are often fully unmethylated while active enhancers show low level of methylation.
- **Gene bodies** are often found to be fully methylated except at enhancer positions.

PMI RESEARCH & DEVELOPMENT

DNA methylation and TF binding

① ES-cell-specific LMRs:
GABPA, ELK1, ELK4, Klf4

② Constitutive LMRs:
CTCF, REST

③ NP-specific LMRs:
Sox9/17, HNF1A/B, Lhx3,
PBX1, Nobox, Pax6

NP : Neuronal progenitors

LMRs are enriched for cell type-specific TF motifs

- DNA cis-regulatory elements (**CRE**), such as promoters and enhancers, are loci that regulate gene expression by functioning as binding sites for transcription factors (**TF**s).

- TF binding can shape DNA methylation locally, and in turn, DNA methylation can influence TF binding leading to a complex loop of interaction involving TFs, chromatin modifying enzymes and chromatin structure.

Stadler et al. Nature 2011 DNA-binding factors shape the mouse methylome at distal regulatory regions

PMI RESEARCH & DEVELOPMENT

Outline

- sbv IMPROVER concept

- Computational challenges open to the community

- The science in the challenge

- **The sbv IMPROVER Israel Epigenomics challenge**

PMI RESEARCH & DEVELOPMENT

The challenge in a glance

The aim of this sbv IMPROVER Challenge is to apply computational approaches to assess the impact of tobacco smoke or aerosol in large methylome datasets obtained from rodent inhalation studies.
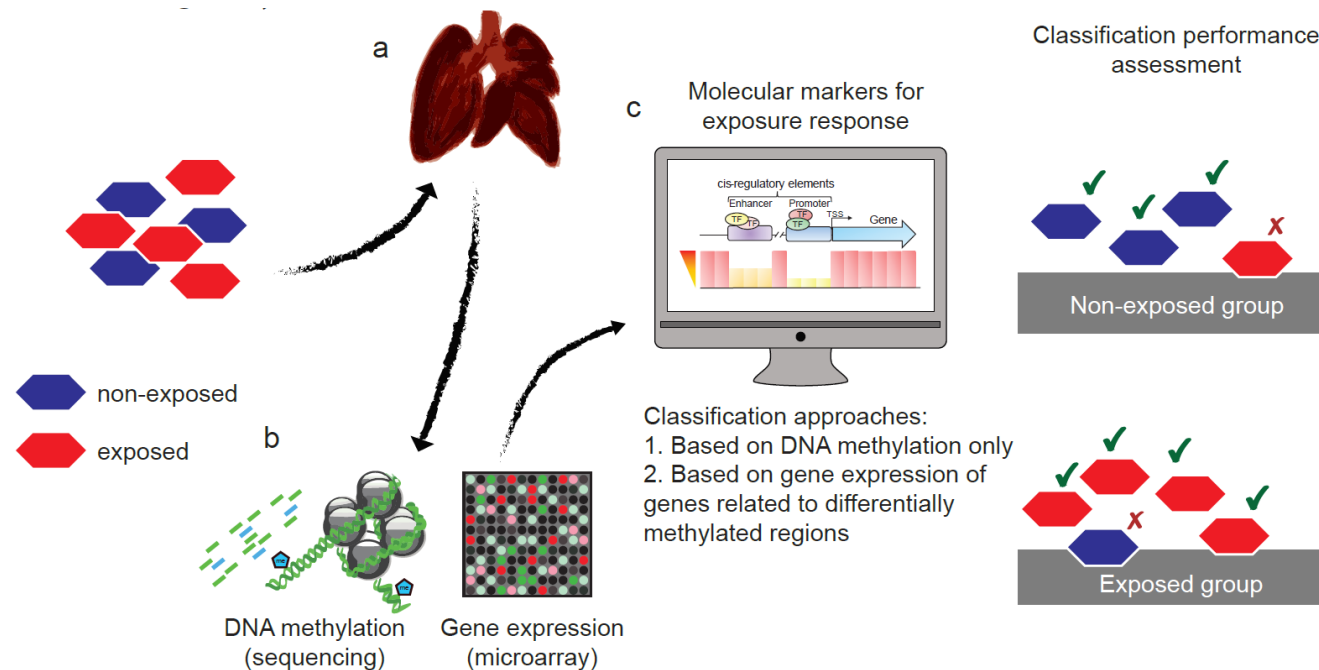

a. Samples of lung tissue were collected from mice belonging to the exposed or non-exposed groups.
b. Gene expression profiles (GEX) and DNA methylation (DNA-Meth) were measured using microarray based technology and Illumina sequencing by synthesis technology, respectively.
c. Participants are provided with GEX and DNA-Meth and asked to develop a classification approach capable of associating subjects to the correct exposure group

PMI RESEARCH & DEVELOPMENT

Data from 3 *in vivo* studies will be used for the challenge

PMI RESEARCH & DEVELOPMENT
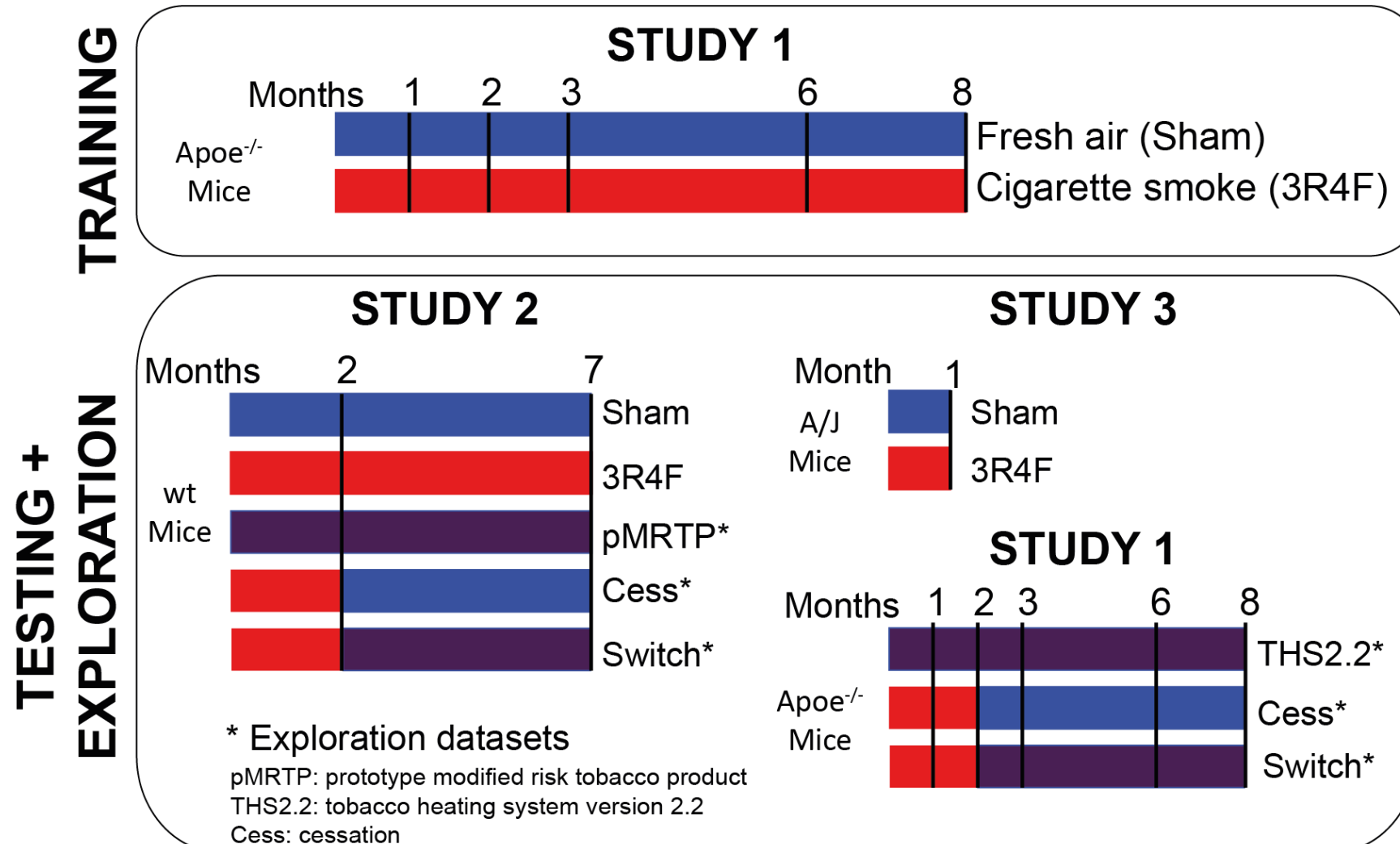
# Smoke exposure signature from differentially methylated cis-regulatory elements

- **Question 1:** *Can a smoke exposure signature be extracted from DNA methylation levels of DNA cis-regulatory elements?*

  - Identify **differentially methylated cis-regulatory elements (CREs),** including annotated (e.g. promoters) and unannotated (e.g. enhancers, insulators…) elements between smoke and fresh air exposed samples

  - Identify **transcription factors** potentially regulating the activity of the differentially methylated CREs

  - Extract a **smoke exposure signature** from DNA methylation levels of the identified CREs

  - **Classify** each sample in the test set using the CRE smoke exposure signature extracted from DNA methylation data, providing the probability that a sample belongs to the 3R4F exposed group
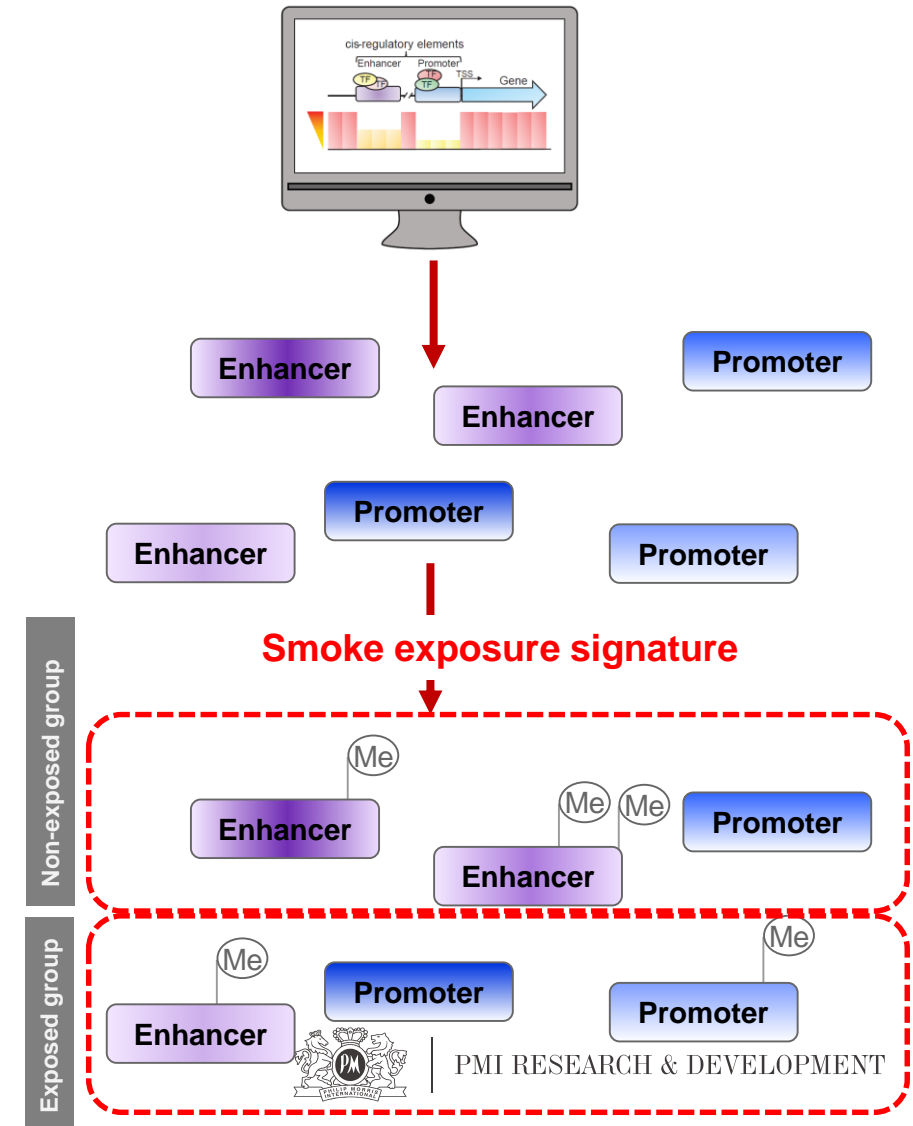
Smoke exposure signature from expression of genes controlled by differentially methylated cis-regulatory elements

- **Question 2:** *Can a smoke exposure signature be extracted from expression data of genes controlled by differentially methylated DNA cis-regulatory elements?*

  – Identify the **target genes** controlled by the CREs from question 1.1

  – Extract a **smoke exposure signature** from the expression data of genes controlled by the 1000 most differentially methylated CREs between smoke and fresh air exposed samples

  – **Classify** each sample in the test set using the smoke exposure signature extracted from expression data, providing the probability that a sample belongs the 3R4F exposed group

PMI RESEARCH & DEVELOPMENT
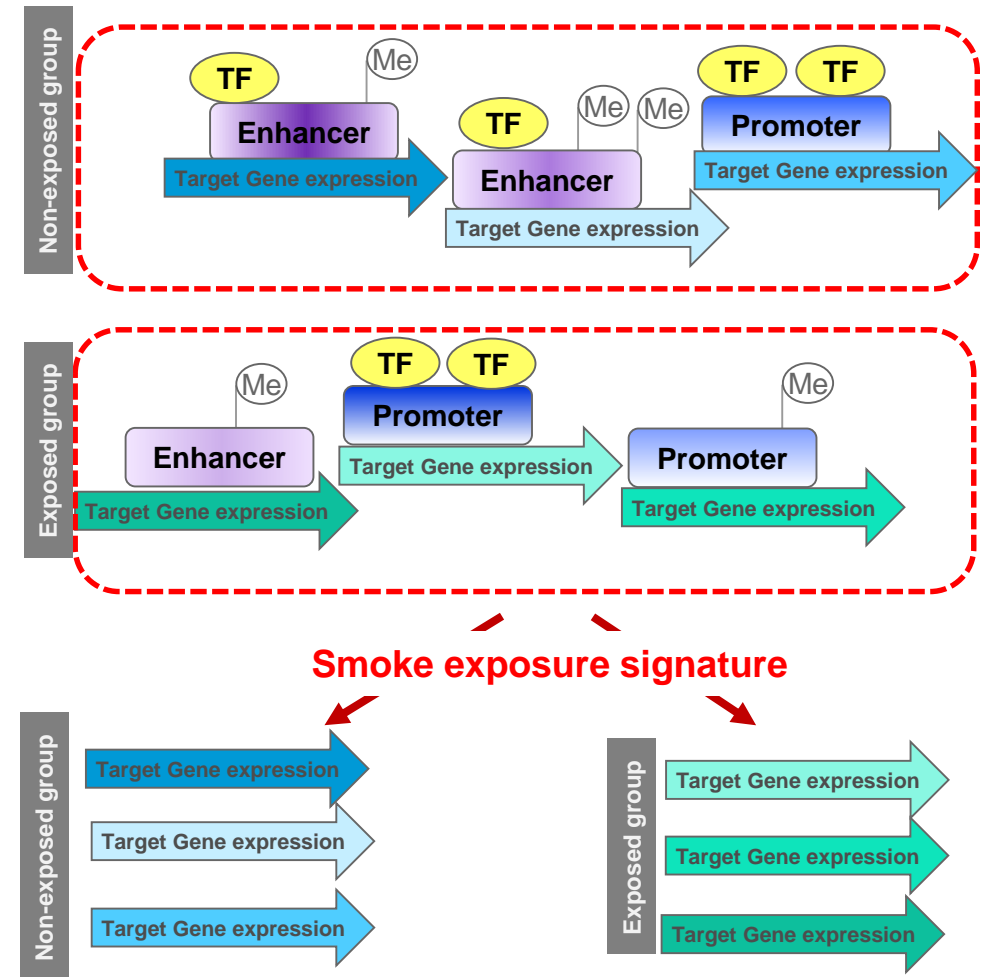
Suggested workflow

Public data bases

ENCYCLOPEDIA OF DNA ELEMENTS

RefSeq

GEO NCBI
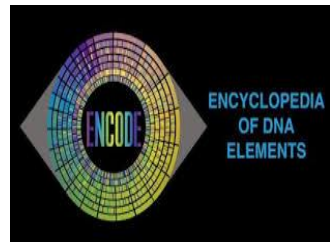
Promoter coordinates
Enhancer marks :
Histone marks
DNAse hypersensitivity (DHS)
Atac seq
….

Intrinsic DNA methylation Signal
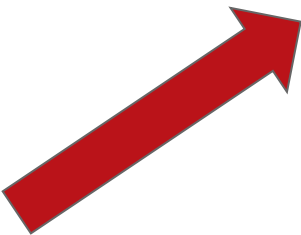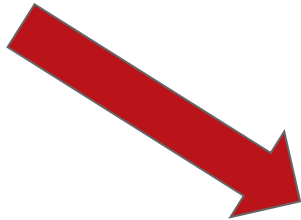
CRE

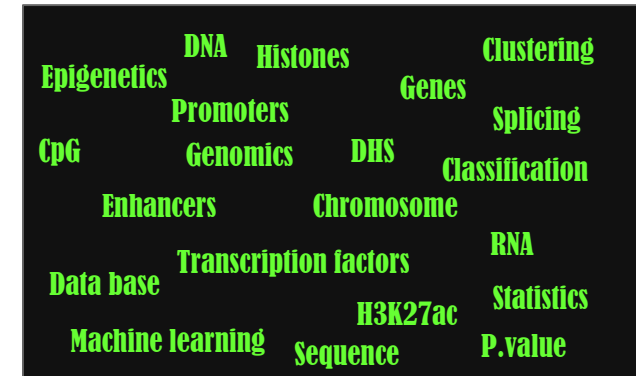chr1 12246 21644
chr3 25694 985673
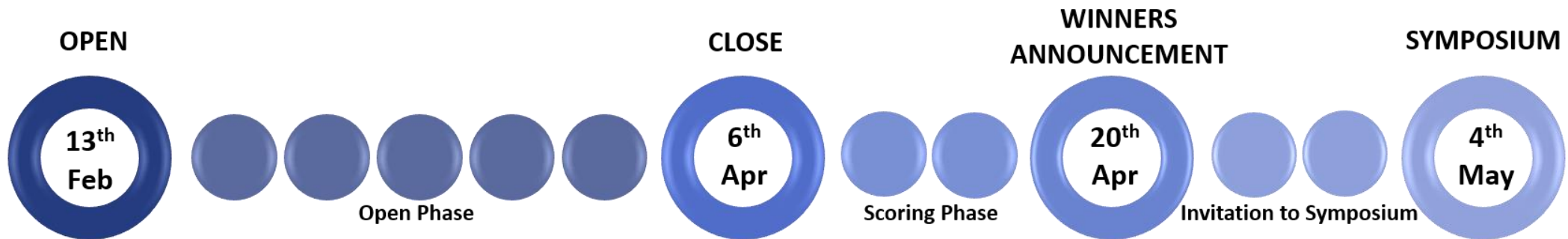chr5 236514 59854
chr6 12946 13958
....
....
....

Quantify methylation level

Cook it !

Epigenetics | DNA | Histones | Clustering | Genes | Promoters | Splicing | CpG | Genomics | DHS | Classification | Enhancers | Chromosome | RNA | Data base | Transcription factors | H3K27ac | Statistics | Machine learning | Sequence | P.value

Timeline

PMI RESEARCH & DEVELOPMENT

Why should you participate?

Have fun!

Win

Win a gift card of:
- USD 1,500 for the 1st team
- USD 1,000 for the 2nd team
- USD 750 for the 3rd team

Top 3 best performing teams will present at the symposium

Enhance your visibility and gain recognition

Receive independent assessment of your methods

Unbiased assessment of methods (known labels)

Large DNA methylation dataset

Access high quality and novel data

Network and engage with peers to advance the field

Join the symposium on 4th May free of charge

Present a poster of your research at the symposium

Present your work at the symposium

PMI RESEARCH & DEVELOPMENT

Challenge ambassador in Israel

Prof. Tamir Tuller

Head of the Laboratory of Computational Systems and Synthetic Biology, Tel Aviv University.

tamirtul@post.tau.ac.il

http://www.cs.tau.ac.il/~tamirtul/

PMI RESEARCH & DEVELOPMENT

Next steps

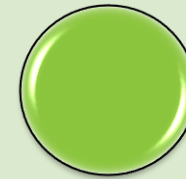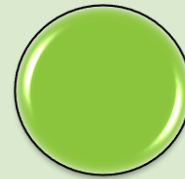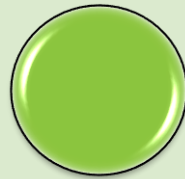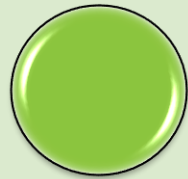## Go to the website

Form a team
(1 to 5
participants)

Get started

Solve the
challenge

Network,
present your
work and enjoy
the symposium
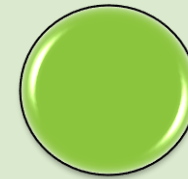
Register

Download
datasets

Upload your
predictions

**It may take
some time**

BEFORE 6th APRIL

PMI RESEARCH & DEVELOPMENT

Get started

Access the data

Participants must register to have access to the data. Upon registration, each registrant will receive an email containing the link to download the data file.
Please note that several gigabytes of data will need to be downloaded and that your time should be planned accordingly.

Submission

Challenge participants are asked to send all files by email to sbvimprover.RD@pmi.com before April 6th, 23h59 CET. If participants teamed up to answer the questions, the team member sending by email the submissions must also include the names and email addresses of each team member (up to max 5 members per team).

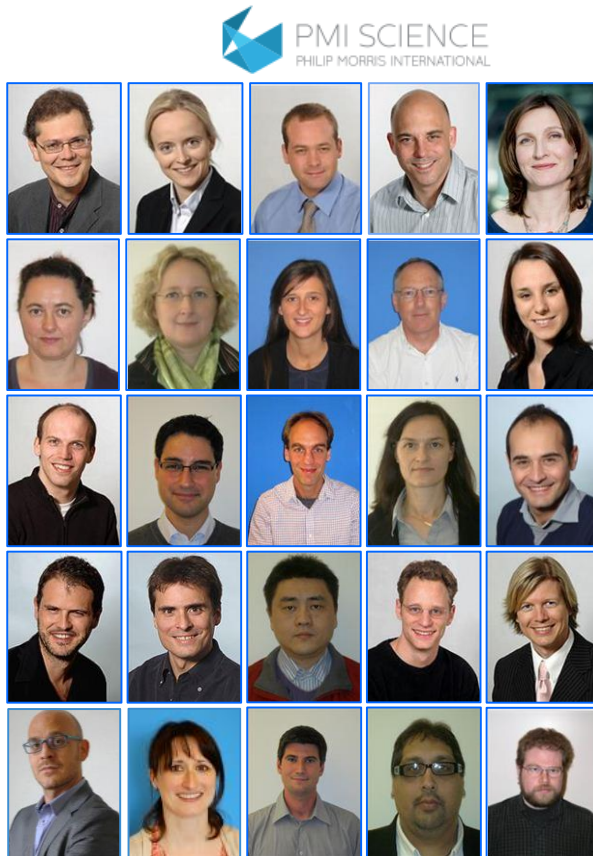PMI RESEARCH & DEVELOPMENT

Symposium 4th May 2017, Tel Aviv

- **Venue:** TBD

- **Agenda:**

9.00 AM	Intro (45 min)	PMI
9.45 AM	Keynote speaker 1 (30 min)	TBD
10.15 AM	Best performer 1 (20 min)	You?
10.35 AM	Coffee break	all
11.00 AM	Keynote speaker 2 (30 min)	TBD
11.30 AM	Best performer 2 (20 min)	You?
11.50 AM	Best performer 3 (20 min)	You?
12.10 PM	Buffet lunch + Networking	all
1.30 PM	Keynote speaker 3 (30 min)	TBD
2.00 PM	Panel discussion (30 min)	all
2.30 PM	Funding and collaboration opportunities with PMI (20 min)	PMI
2.50 PM	Closing	PMI
3.00 PM	Poster session and afternoon cocktail	all

PMI RESEARCH & DEVELOPMENT

Acknowledgements

PMI RESEARCH & DEVELOPMENT