Open Science in Practice
A personal perspective….

Bart Deplancke

Lab Overview: Understanding genome function

"Techno"

"Adipo"

"Geno"

Microfluidics
Single cell analysis

Hens et al., Nature Methods, 2011
Gubelmann et al., Mol Syst Biol, 2013
Simicevic et al., Nature Methods, 2013
Isakova et al., Nature Methods, 2017
Gardeux et al., Bioinformatics, 2017

White and brown adipogenesis
Mesenchymal stem cell function

Raghav et al., Molecular Cell, 2012
Gubelmann et al., eLife, 2014
Pradhan et al., Sci Rep, 2017
Pradhan et al., Obesity, 2017

Human and *Drosophila*
systems genetics

Kiplinen, Waszak* et al., Science, 2013*
Huang, Massouras* et al.,, Genome Res, 2014*
Massouras, Osman* et al., Nat Comm, 2015*
Waszak, Delaneau* et al., Cell, 2015*
Deplancke et al., Cell, 2016

Genomics opened doors

- Before the Human Genome Project, scientists only shared their research findings in scientific journals

- In 1996, those involved in the HGP → *all new information produced should be made freely available to all within 24h* Implemented when Myriad Genetics patented the breast cancer gene *BRCA2* and remove the rights from other scientists to profit from carrying out research on the gene

- The HGP showed how the internet can play a vital part in collective scientific research. Now more scientists are collaborating – and inviting amateurs and colleagues from other disciplines to get involved

Genetic Testing for Breast Cancer

BRCA1 BRCA2 — BReast Cancer genes 1 and 2. Having a mutation in either of these genes increases a woman's risk for breast and ovarian cancer

45% TO 65% — Of women who inherit a mutation will develop Breast Cancer by age 70 years

Genomics opened doors: for our lab

Primer-initiated sequence synthesis to detect and assemble structural variants

Andreas Massouras[1,5], Korneel Hens[1,5], Carine Gubelmann[1], Swapna Uplekar[2], Frederik Decouttere[3], Jacques Rougemont[4], Stewart T Cole[2] & Bart Deplancke[1]

To assess the false positive and negative rate of variant detection, we first analyzed **36 bp Illumina single-ended reads from the Salmonella paratyphi A AKU12601 genome (21-fold coverage).** We used Maq and PrInSeS using the actual genome as reference. Then we did the same analysis again, but using strain **ATCC9150 as reference**.

For the Drosophila melanogaster whole genome dataset, we selected RAL-304, a line out of the 15 for which both single- and paired-end data has been released. **We obtained single-end reads (15-fold coverage) from http://www.ncbi.nlm.nih.gov/sra?term=RAL-304 and paired-end reads from http://www.hgsc.bcm.tmc.edu/projects/dgrp/lines/304/Illumina/ (also 15-fold coverage).**

Genomics opened doors: to others

Molecular Cell
Article

Integrative Genomics Identifies the Corepressor SMRT as a Gatekeeper of Adipogenesis through the Transcription Factors C/EBPβ and KAISO

Sunil K. Raghav,[1,4] Sebastian M. Waszak,[1,4] Irina Krier,[1] Carine Gubelmann,[1] Alina Isakova,[1] Tarjei S. Mikkelsen,[2,3] and Bart Deplancke[1,*]

[1]Laboratory of Systems Biology and Genetics, Institute of Bioengineering, School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland
[2]Broad Institute, 7 Cambridge Center, Cambridge, MA 02142, USA
[3]Harvard Stem Cell Institute and Department of Stem Cell and Regenerative Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA 02138, USA
[4]These authors contributed equally to this work
*Correspondence: bart.deplancke@epfl.ch
DOI 10.1016/j.molcel.2012.03.017

SMRT, NCoR1, and RNA pol II ChIP-Seq data sets
associated with this study are available at ArrayExpress
under the accession number E-MTAB-1031.

E-MTAB-1031 - ChIP-seq of mouse embryonic fibroblast-adipose like cell line 3T3-L1 to identify binding sites of NCoR1 and SMRT following induction of differentiation, and RNA Pol-II after SMRT knock down

Status	*Released on 23 April 2012, last updated on 3 May 2014*
Organism	Mus musculus
Samples (10)	Click for detailed sample information and links to data
Protocols (6)	Click for detailed protocol information
Description	ChIP-seq of mouse embryonic fibroblast-adipose like cell line 3T3-L1 to identify binding sites of NCoR1 and SMRT following induction of differentiation, and RNA Pol-II after SMRT knock down
Experiment types	ChIP-seq, ChiP-seq, development or differentiation, in vitro
Contacts	✉ Irina Krier <irina.krier@epfl.ch>, ✉ Bart Deplancke <bart.deplancke@epfl.ch>
Citation	Integrative genomics identifies the co-repressor SMRT as a gatekeeper of adipogenesis through the transcription factors C/EBP? and KAISO. Sunil K. Raghav, Sebastian M. Waszak, Irina Krier, Carine Gubelmann, Alina Isakova, Tarjei S. Mikkelsen, and Bart Deplancke.
MINSEQE	✱ Exp. design ✱ Protocols ✱ Variables ─ Processed ✱ Seq. reads
Files	Investigation description ⬇ E-MTAB-1031.idf.txt Sample and data relationship ⬇ E-MTAB-1031.sdrf.txt

Integrative genomics identifies the corepressor SMRT as a gatekeeper of adipogenesis through the transcription factors C/EBPβ and KAISO

SK Raghav, SM Waszak, I Krier, C Gubelmann, A Isakova, TS Mikkelsen, ...

Molecular cell 46 (3), 335-350

[HTML] ZBTB33 binds unmethylated regions of the genome associated with actively expressed genes

A Blattler, L Yao, Y Wang, Z Ye… - Epigenetics & …, 2013 - epigeneticsandchromatin. …

Background DNA methylation and repressive histone modifications cooperate to silence promoters. One mechanism by which regions of methylated DNA could acquire repressive histone modifications is via methyl DNA-binding transcription factors. The zinc finger protein

☆ 🔗 Cited by 31   Related articles   All 15 versions   Web of Science: 21   »

[HTML] Genome-wide map of nuclear protein degradation shows NCoR1 turnover as a key to mitochondrial gene regulation

A Catic, CY Suh, CT Hill, L Daheron, T Henkel… - Cell, 2013 - Elsevier

Summary Transcription factor activity and turnover are functionally linked, but the global patterns by which DNA-bound regulators are eliminated remain poorly understood. We established an assay to define the chromosomal location of DNA-associated proteins that

☆ 🔗 Cited by 23   Related articles   All 26 versions   Web of Science: 19   »

Quantification of cooperativity in heterodimer-DNA binding improves the accuracy of binding specificity models

A Isakova, Y Berset, V Hatzimanikatis… - Journal of Biological …, 2016 - ASBMB

Abstract Many transcription factors (TFs) have the ability to cooperate on DNA elements as heterodimers. Despite the significance of TF heterodimerization for gene regulation, a quantitative understanding of cooperativity between various TF dimer partners and its impact

☆ 🔗 Cited by 10   Related articles   All 8 versions   Web of Science: 7   »

### ATF4 licenses C/EBPβ activity in human mesenchymal stem cells primed for adipogenesis

DM Cohen, KJ Won, N Nguyen, MA Lazar, CS Chen… - Elife, 2015 - elifesciences.org

Abstract A well-established cascade of transcription factor (TF) activity orchestrates adipogenesis in response to chemical cues, yet how cell-intrinsic determinants of differentiation such as cell shape and/or seeding density inform this transcriptional program

☆ 🔗 Cited by 7   Related articles   All 13 versions   Web of Science: 4   »

Academic Altruism : a Gateway to Open Science / Source
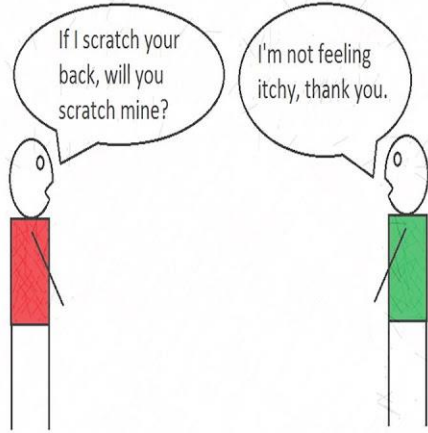
Many of our analyses have benefited from the altruistic efforts of our colleagues to develop powerful analysis packages

Bowtie	Uses a Burrows-Wheeler transform to create a permanent, reusable index of the genome; 1.3 GB memory footprint for human genome. Aligns more than 25 million Illumina reads in 1 CPU hour. Supports Maq-like and SOAP-like alignment policies	Yes	Yes	No	Yes, POSIX Threads	Free, Artistic	link	[32]	2009
HIVE-hexagon	Uses a hash table and bloom matrix to create and filter potential positions on the genome. For higher efficiency uses cross-similarity between short reads and avoids realigning non unique redundant sequences. It is faster than bowtie and bwa and allows indels and divergent sensitive alignments on viruses, bacteria, and more conservative eukaryotic alignments.	Yes	Yes	Yes	Yes	Proprietary, freeware for academic and noncommercial users registered to HIVE deployment instance	link	[33]	2014
BWA	Uses a Burrows-Wheeler transform to create an index of the genome. It's a bit slower than bowtie but allows indels in alignment.	Yes	Low quality bases trimming	Yes	Yes	Free, GPL	link	[34]	2009

Academic Altruism : a Gateway to Open Science / Source

Software

Check our github for the latest updates on our software.

Check our youtube channel for the latest tutorials on our software.

- ASAP
  ASAP: a Web-based platform for the analysis and inter-active visualization of single-cell RNA-seq data (Gardeux et al. Bioinformatics, 2017)

- Get Prime
  Gene- or transcript-specific primer design for real-time PCR (Gubelmann et al., Database, 2012)

- PrInSeS
  Primer-initiated de novo sequence synthesis for clone identification (cDNA screens), validation, and genomic annotation (indels) (Massouras et al., Nature Methods, 2010)

- ABS
  Detection of amplification-biased sites in allele-specific analysis of ChIP-seq data (Waszak et al., Bioinformatics, 2014)

- TIDY
  "Transcription factor-DNA Interaction Detection in

GETPrime 2.0: gene- and transcript-specific qPCR primers for 13 species including polymorphisms

Fabrice P.A. David[1,2], Jacques Rougemont[1,2,*] and Bart Deplancke[2,3,*]

NEWS: Last update finished on 2015 November 12 using Ensembl release 81. More species are available (13, see below).

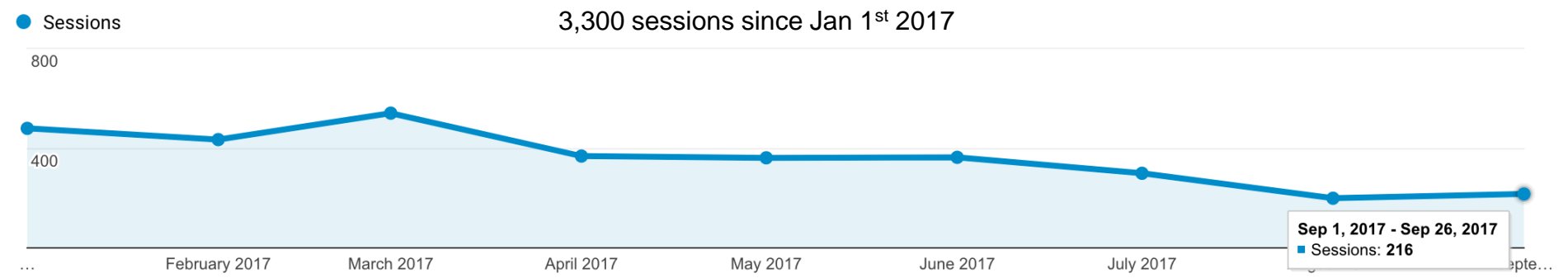Ensembl release [81]
Organism [Homo sapiens]
Limit [100]

Comma or space-separated list of identifiers:

pparg

Search

45 primer pairs found. Download

Search all columns: [_____]

ID ▲	Gene	Transcripts	Rank	# Transcripts	Amplicon length	Forward primer	Tm fwd	Reverse primer	Tm rev	Ensembl status
1909238	ENSG00000132170 PPARG	ENST00000397010 ENST00000397029 ENST00000309576 ENST00000397015 ENST00000397012 ENST00000397026 ENST00000497594 ENST00000438682 ENST00000397000 ENST00000477039 ENST00000287820 ENST00000397023 ENST00000396999	★★★ (1)	13/14	80	3:12379863-12379884 TTCCATTCACAAGAACAGATCC 2 SNPs	58	3:12379922-12381333 CTTTGATTGCACTTTGGTACTC 4 SNPs	58	Known

● Sessions

3,300 sessions since Jan 1ˢᵗ 2017

800

400

February 2017   March 2017   April 2017   May 2017   June 2017   July 2017

Sep 1, 2017 - Sep 26, 2017
■ Sessions: 216

Academic Altruism : a Gateway to Open Science / Source

Software

Check our github for the latest updates on our software.

Check our youtube channel for the latest tutorials on our software.

- **ASAP**
  ASAP: a Web-based platform for the analysis and inter-active visualization of single-cell RNA-seq data (Gardeux et al. Bioinformatics, 2017)

- **Get Prime**
  Gene- or transcript-specific primer design for real-time PCR (Gubelmann et al., Database, 2012)

- **PrInSeS**
  Primer-initiated de novo sequence synthesis for clone identification (cDNA screens), validation, and genomic annotation (indels) (Massouras et al., Nature Methods, 2010)

- **ABS**
  Detection of amplification-biased sites in allele-specific analysis of ChIP-seq data (Waszak et al., Bioinformatics, 2014)

- **TIDY**
  "Transcription factor-DNA Interaction Detection in

Explore GitHub

Science

Scientists around the world are working together to solve some of the biggest questions in research. Take a look at some of the examples featured here to find out more.

Deplancke's Lab
DeplanckeLab

Laboratory of Systems Biology And Genetics (LSBG) held by B. Deplancke at EPFL

Overview	Repositories 3	Stars 0	Followers 1	Following 0

Popular repositories

ASAP
ASAP : Automated Single-cell Analysis Pipeline
● R ★ 3 ⑂ 1

ASAP-1
Forked from NHLBI-BCB/ASAP
ASAP : Automated Single-cell Analysis Pipeline
● Java

SmartSeq2

ASAP: Automated Single-cell Analysis Pipeline

Existing software:

SINCERA, MAST, PAGODA, SCell, FastProject…

Disadvantages:

- Restricted set of algorithms & methods

- Lack of interactivity & visualization

- Required knowledge of programming and/or statistics

- Non-standardized input

- Difficult installation of libraries & dependencies

- Depends on personal computer processing power

Dr. Petra Schwalie

Dr. Vincent Gardeux

Adrian Shajkofci
(Master's student (EPFL))

ASAP: Automated Single-cell Analysis Pipeline
https://asap.epfl.ch/

- Combines state-of-the art open source, single-cell specific algorithms written in R, Python, & Java

- **Interactive** & user-friendly web interface with 2D/3D visualization

- Robust parser for a wide range of input data

- Nothing to install for the end user

- Centralized computational resources

- Multi-user

Gene expression

ASAP: a web-based platform for the analysis and interactive visualization of single-cell RNA-seq data

Vincent Gardeux[1,2], Fabrice P. A. David[2,3], Adrian Shajkofci[1], Petra C. Schwalie[1,2] and Bart Deplancke[1,2,*]

⇒ **What took considerable time for a skilled bioinformatician can now be achieved with ASAP in few minutes without any prior bioinformatics knowledge**

ASAP : Automated Single-cell Analysis Pipeline

Read-only mode Clone to get full control

Project 'Dueck 2015' on mouse

| Parsing ✓ | Filtering ✓ | Normalization ✓ | Visualization ✓ | Cell clusters 4 | Differential expression 4 | Gene enrichment 3 |

PCA ✓   tSNE ✓   MDS ✓   ZIFA ✓   Heatmap ✓   Correlation ✓

X Dim1 **Y** Dim2 **Z** Dim3 **Color by** Gene expression **Gene identifier** [   ] **Dataset** Normalized data [Plot]

2d plot

Full Screen | Toggle legend | Toggle cell names

3d plot

Full Screen | Toggle legend | Toggle cell names

Search
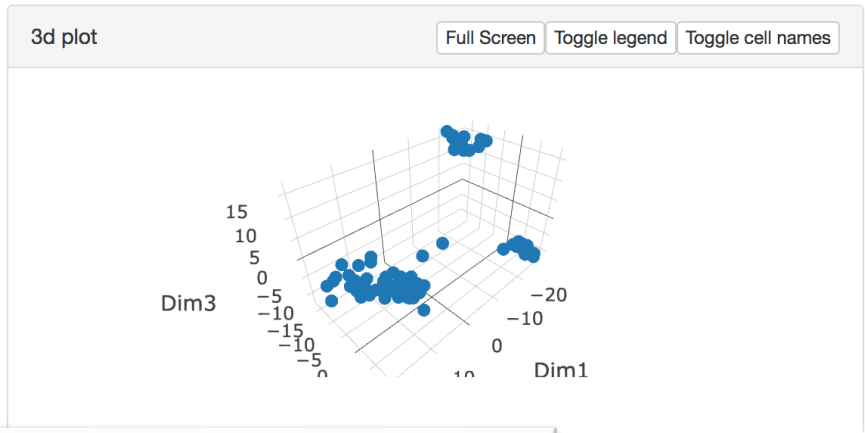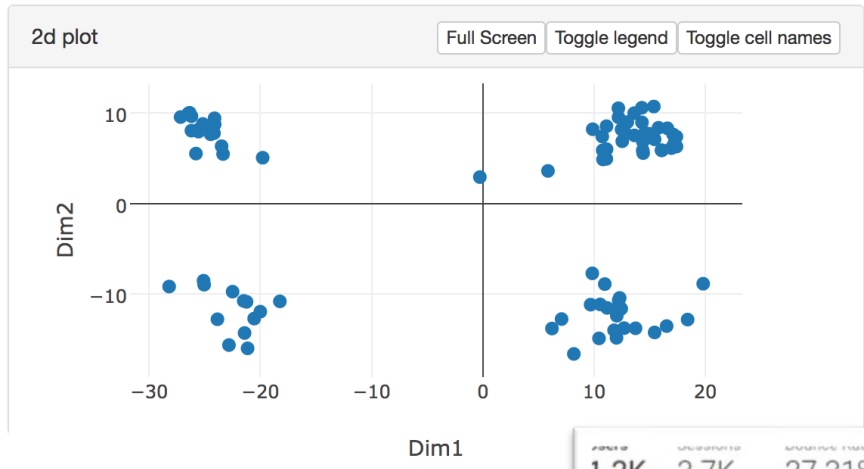
New Results

View current version of this article

ASAP: a Web-based platform for the analysis and inter-active visualization of single-cell RNA-seq data

Download PDF   Share

Vincent Gardeux, Fabrice David, Adrian Shajkofci, Petra C Schwalie, Bart Deplancke
doi: https://doi.org/10.1101/096222

Email   Citation Tools

Tweet   Like 0   G+

- The community gets to immediately contemplate your work (collect citations!)

ASAP: a Web-based platform for the analysis and inter-active ... - bioRxiv
www.biorxiv.org/content/early/2016/12/22/096222 ▾
by V Gardeux - 2016 - Cited by 1 - Related articles
Dec 22, 2016 - **ASAP**: a Web-based platform for the analysis and inter-active visualization of single-cell RNA-seq data. Vincent Gardeux, Fabrice David, Adrian ...

- The community gets to immediately contemplate your work (collect citations!)

- Lots of online attention triggers journal interest (they actively scan for interesting papers)

- It puts a date stamp on your work / discovery claim

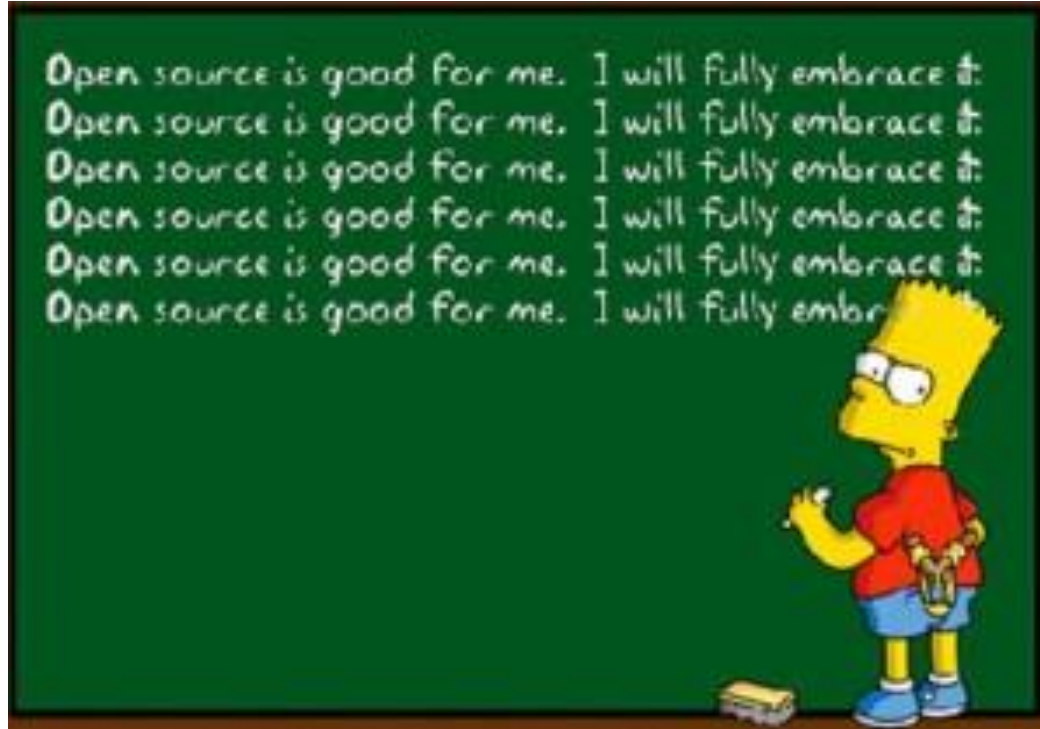- Journals are increasingly coming to peace with BioRxiv papers

Open Source, yes, but….

There's a religious war between the defenders of open-source and commercial product supporters.

The clinical / omic data deluge

How to manage this in the lab, the clinic, the diagnostic company?

The journey from Genome to Genohm

Disclaimer: I am co-founder, shareholder, and current board member

The SLims Platform
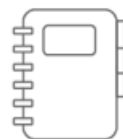
3 Offices – 40 FTE

Gent

Lausanne

Durham

Why pay for your product?
Why not open source?

- That open source software is free is a myth
- No shame in making money for providing great value & service, or spending money on software to help grow your business / organization
- Commercial software is typically well suited to support systematic, "boring" tasks, open source for more expert, specialized tasks
- Open source is not a philosophy, nor a business model, it's a software feature
  - → depends on what you're looking for
  - → commercial entities provide the resources to develop the code, provide quality assurance and deliver professional support and maintenance
  - → ISO/TC 276 Biotechnology & The **General Data Protection Regulation** (**GDPR**) (Regulation (EU) 2016/679)

Questions?