# Bioinformatics Consulting at HMS

## The Harvard Chan Bioinformatics Core

*Department of Biostatistics, Harvard Chan School of Public Health, Boston, MA, USA*

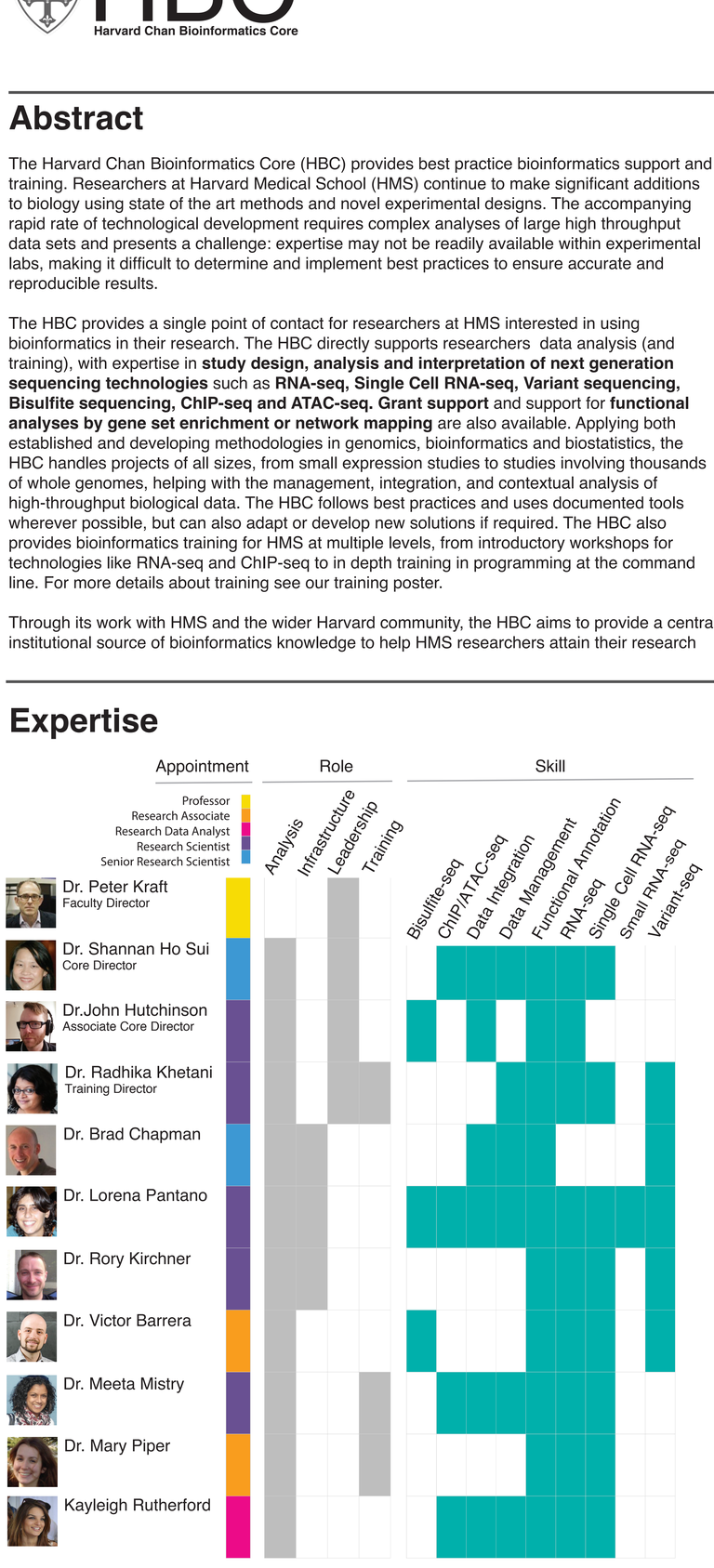
## Abstract

The Harvard Chan Bioinformatics Core (HBC) provides best practice bioinformatics support and training. Researchers at Harvard Medical School (HMS) continue to make significant additions to biology using state of the art methods and novel experimental designs. The accompanying rapid rate of technological development requires complex analyses of large high throughput data sets and presents a challenge: expertise may not be readily available within experimental labs, making it difficult to determine and implement best practices to ensure accurate and reproducible results.

The HBC provides a single point of contact for researchers at HMS interested in using bioinformatics in their research. The HBC directly supports researchers data analysis (and training), with expertise in **study design, analysis and interpretation of next generation sequencing technologies** such as **RNA-seq, Single Cell RNA-seq, Variant sequencing, Bisulfite sequencing, ChIP-seq and ATAC-seq. Grant support** and support for **functional analyses by gene set enrichment or network mapping** are also available. Applying both established and developing methodologies in genomics, bioinformatics and biostatistics, the HBC handles projects of all sizes, from small expression studies to studies involving thousands of whole genomes, helping with the management, integration, and contextual analysis of high-throughput biological data. The HBC follows best practices and uses documented tools wherever possible, but can also adapt or develop new solutions if required. The HBC also provides bioinformatics training for HMS at multiple levels, from introductory workshops for technologies like RNA-seq and ChIP-seq to in depth training in programming at the command line. For more details about training see our training poster.

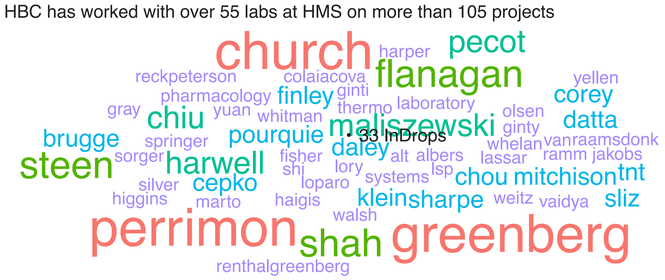
Through its work with HMS and the wider Harvard community, the HBC aims to provide a central institutional source of bioinformatics knowledge to help HMS researchers attain their research

## Expertise

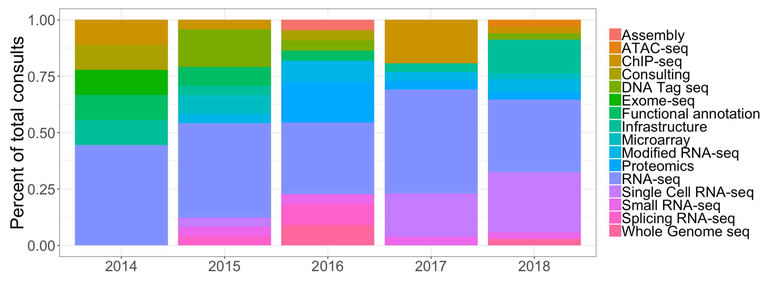



## Projects

- the HBC has worked with over 55 labs at HMS on more than 105 projects




- analysis of gene expression by RNA-seq remains the most common bioinformatics need
- we continue to see large increases in demand for Single Cell RNA-seq, including Nuc-seq
  - 33 InDrops consultations (13 experimental design, 20 analyses)
  - 10 10X Chromium consultations



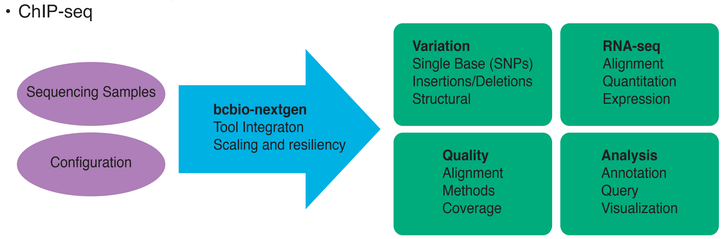

## Infrastructure

### Pipelines

bcbio https://bcbio-nextgen.readthedocs.io

**Approach**
- Reproducible, Scaleable, Automated, Documented, Self-contained, Interoperable
- Open Source and Community Driven

**Functions**
- Variant Calling (exome, whole genome, structural, CNVs, cancer)
- RNA-seq (bulk, single cell, small RNAs)
- ChIP-seq




### R Packages

DEGreport


DOI: 10.18129/B9.bioc.DEGreport

Report of DEG analysis

Bioconductor version: Release (3.6)

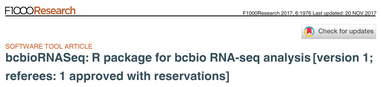
Creation of a HTML report of differential expression analyses of count data. It integrates some of the code mentioned in DESeq2 and edgeR vignettes, and report a ranked list of genes according to the fold changes mean and variability for each selected gene.

Author: Lorena Pantano [aut, cre], John Hutchinson [ctb], Victor Barrera [ctb], Mary Piper [ctb], Kenneth Daily [ctb], Thanneer Malai Perumal [ctb], Rory Kirchner [ctb], Michael Steinbaugh [ctb]

Maintainer: Lorena Pantano <lorena.pantano at gmail.com>

Citation (from within R, enter `citation("DEGreport")`):

Pantano L (2017). *DEGreport: Report of DEG analysis.* R package version 1.14.0.




SOFTWARE TOOL ARTICLE

**bcbioRNASeq: R package for bcbio RNA-seq analysis [version 1; referees: 1 approved with reservations]**

Michael J. Steinbaugh[1*], Lorena Pantano[1*], Rory D. Kirchner[1], Victor Barrera[1], Brad A. Chapman[1], Mary E. Piper[1], Meeta Mistry[1], Radhika S. Khetani[1], Kayleigh D. Rutherford[1], Oliver Hofmann[2], John N. Hutchinson[1], Shannan Ho Sui[1]

[1]Harvard T.H. Chan School of Public Health, Boston, MA, 02115, USA
[2]University of Melbourne Center for Cancer Research, Melbourne, VIC, 3000, Australia

[*] Equal contributors

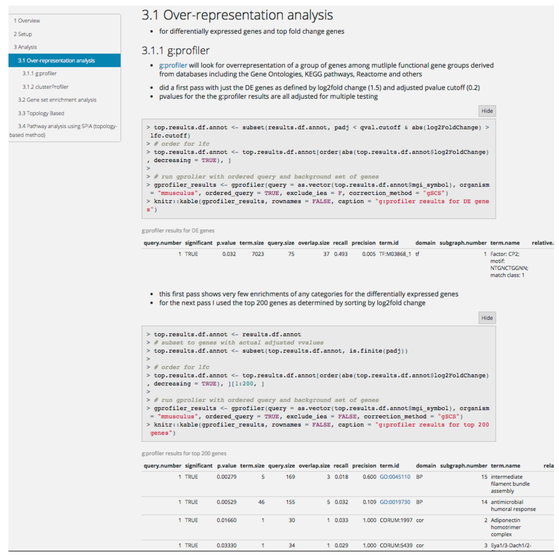
## Working With Us

- contact us early for help with experimental design
- schedule an initial meeting, free of charge
- we will create a timeline with deliverables
- we will provide a quote covering personnel, data storage and compute costs
- ***subsidized rates are available for HMS researchers on the quad***
- progress is regularly documented on a secure project site
- all data sets, results and documentation are shared

### Programs

```
bamtools,2.4.0
bcbio-nextgen,0.9.8a0-8183767
bcbio-variation,0.2.6
bcftools,1.3
bedtools,2.24.0
biobambam,2.0.42
bioconductor-bubbletree,2.1.5
bowtie2,2.2.8
bwa,0.7.13
chanjo,
cnvkit,0.7.11
cufflinks,2.2.1
cutadapt,1.9.1
fastqc,0.11.5
featurecounts,1.4.4
freebayes,1.0.2
gatk,3.2-2-gec30cee
gatk-framework,3.5.21
gemini,0.18.3
grabix,0.1.6
hisat2,2.0.3beta
htseq,0.6.1p1
lumpy-sv,0.2.12
manta,0.29.6
metasv,0.4.0
mutect,1.1.5
novoalign,3.04.04
novosort,V3.00.02
oncofuse,1.1.0
phylowgs,20150714
picard,1.141
platypus-variant,0.8.1
qualimap,2.1.3
rna-star,2.4.1d
rtg-tools,3.6
sailfish,0.9.0
salmon,0.6.0
sambamba,0.6.1
samblaster,0.1.22
samtools,1.3.1
scalpel,0.5.1
snpeff,4.2
vardict,2016.02.19
vardict-java,1.4.5
variant-effect-predictor,83
varscan,2.4.1
vcflib,1.0.0_rc0
vt,2015.11.10
wham,1.7.0.162
```

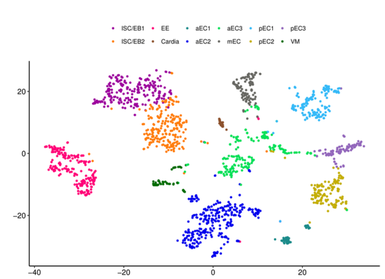
### Rmarkdown report with code




- we provide methods, publication quality figures and help with GEO submissions

## Example Results

Ruei-Jiun Hung, Yanhui Hu, **Rory Kirchner**, Fangge Li, Chiwei Xu, Aram Comjean, Sudhir Gopal Tattikota, Wei Roc Song, **Shannan Ho Sui**, Norbert Perrimon. (2018). A cell atlas of the adult Drosophila midgut. *bioRxiv* (doi: https://doi.org/10.1101/410423)




**T-SNE visualization of 1753 adult fly intestinal gut cells:** ISC/EB, intestinal stem cells/ enteroblast; EE, enteroendocrine; aEC, anterior enterocyte; mEC, middle enterocyte; pEC, posterior enterocyte.


**Expression levels and percentage of cells expressing the top 3 markers in each cluster as a dot plot**

**Contact us:**
We are located at the Harvard School of Public Health, SPH2, 2nd floor, Room 215.
Projects: bioinformatics@hsph.harvard.edu
Training: hbctraining@hsph.harvard.edu (see our other poster for details!)
Website : bioinformatics.hms.harvard.edu