Wardrobe experiment management system for integrated analysis of epigenomics data

Andrey V. Kartashov, M.S.[1] and Artem Barski, Ph.D. [1,2]

[1]Division of Allergy and Immunology, [2]Division of Human Genetics, Cincinnati Children's Hospital Medical Center and Department of Pediatrics, College of Medicine, University of Cincinnati, Cincinnati, OH

Correspondence: Artem.Barski@cchmc.org

Abstract

High-throughput sequencing has revolutionized biology by enhancing our ability to perform genome-wide studies. However, due to lack of bioinformatics expertise, technologies are still beyond the capabilities of many laboratories. Herein, we present Wardrobe server, which allows users to store, visualize and analyze epigenomics and transcriptomics data using a biologist-friendly GUI without the need for programming expertise. Predefined pipelines allow users to download data from core facilities or public databases, visualize data on a genome browser, calculate RPKMs and identify islands. Advanced capabilities include differential gene expression and binding analysis, and creation of average tag density profiles and heatmaps. Wardrobe is available at https://code.google.com/p/genome-tools/.

Keywords

Epigenomics, transcriptomics, data analysis, bioinformatics, ChIP-seq, RNA-Seq, DNase-Seq, ATAC-Seq

Introduction

The recent proliferation of next-generation sequencing (NGS)-based methods for analysis of gene expression, chromatin structure and protein-DNA interactions has opened new horizons for molecular biology. These methods include RNA sequencing (RNA-Seq)[1], chromatin immunoprecipitation sequencing (ChIP-Seq)[2], DNase I hypersensitive sites sequencing (DNase-Seq)[3], micrococcal nuclease sequencing (MNase-Seq)[4], global run-on sequencing (GRO-Seq)[5], assay for transposase-accessible chromatin sequencing (ATAC-Seq)[6], and others. On the “wet lab” side, these methods are more or less well established and can be performed by experienced molecular biologists; however, analysis of the resulting data requires computational expertise that many molecular biologists do not possess or have access to. Another difficulty is re-utilization of published datasets: although authors comply with the longstanding requirement to deposit raw data files into databases such as Sequence Read Archive (SRA) or Gene Expression Omnibus (GEO), it is impossible to analyze these datasets without special expertise. Even when processed data files (e.g. gene expression values) are available, direct comparison between datasets is ill advised because different laboratories use different pipelines (or different software versions). This situation means that biologists require the help of bioinformaticians even for simple tasks, such as viewing their own data on a genome browser, and thus places these exciting techniques beyond the reach of the majority of scientists. Even when bioinformaticians are available, differences in priorities can result in delays and misunderstandings that are damaging to the research effort. An optimal way to mitigate these problems is by creating user-friendly tools that can help biologists perform at least some of the tasks without the help of bioinformaticians.

Several stand-alone programs and web services are available for the analysis of NGS data. However, the majority of currently available tools are command line or R based and only a few tools have graphical user interfaces that do not require Linux or programming skills. The programs Genespring [7], Partek [8] and Golden Helix [9] can be run on regular desktop computers and allow analysis of gene expression or genetic variation. However, users have to load the data manually and store it on their desktop computers; given the sheer volume of NGS datasets, this makes data analysis complicated at best. Furthermore, these tools do not allow for seamless integration of multiple, published or locally produced datasets. Though Illumina Basespace [10] and Galaxy server [11] allow for both storage and analysis of data and have integrated viewing tools. However, they require transfer of data outside the institution (which may be prohibited by HIPAA regulations in some cases) and provide only limited storage space for user data. Although Galaxy provides the opportunity to run analysis without command lines, users still need to select parameters each time, which requires a deep understanding of each tool. Absence of highly formalized pipelines may result in inexperienced users comparing “apples to oranges”. In summary, few of the available tools provide biologist-friendly interfaces, and none integrate data storage, display and analysis with a user-friendly interface.

We therefore developed Wardrobe, a biologist-friendly solution for integrated acquisition, storage, display and analysis of NGS data, primarily for use in epigenomics research. Wardrobe allows for download of raw data from core facilities or databases (e.g. GEO), read mapping and display on an integrated mirror of the University of California, Santa Cruz (UCSC) genome browser [12], quality control and basic and advanced data analysis. The automated basic analysis pipeline incorporates both widely used tools developed by others (e.g. Bowtie[13], STAR [14], FASTX [15] and MACS [16]) and several of our own programs. Customizable advanced analysis includes tools for comparing gene expression DESeq [17] and ChIP-Seq (MAnorm [18]) profiles between samples or groups of samples and creating gene lists, average tag density profiles and heatmaps. Most of the data are precomputed and stored in an SQL database and can be accessed via a convenient web interface by biologists. Bioinformaticians can also access the data via Python or R interfaces on the server or even on the web via third-party tools like RStudio. Wardrobe can be run on Linux or MacOSX systems with attached storage. Source code and installation instructions are available at https://code.google.com/p/genome-tools/. A limited-functionality demo version is available at http://demo.wardrobe.porter.st/ems.

System overview

Wardrobe allows users to upload, store and analyze NGS data. The workflow consists of two parts: basic and advanced analysis (Fig. 1). The basic analysis includes operations that do not require comparison of samples: data download, quality control, calculation of RPKMs (reads per kilobase of transcript per million reads mapped), island identification and upload to a built-in mirror of the UCSC genome browser. Advanced analysis includes comparing gene expression or ChIP-Seq profiles between samples. A flexible data ownership system is implemented: while all users can see all datasets on a local mirror of the UCSC genome browser, only members of the laboratories that own the data can access and analyze datasets within Wardrobe web interface or download it. Laboratory-level administrators can elect to share data with other laboratories. However, trusted bioinformaticians can have access to all datasets outside of web-interface—e.g. via RStudio. We believe that this setup strikes a balance between maintaining data ownership and encouraging collaborations.

Basic analysis

Basic analysis includes operations that are performed on a single library (Fig. 1B). Analysis starts by entering the experiment description into Wardrobe. This information will be used to select the appropriate genome and analysis pipeline. Raw data can be directly downloaded by Wardrobe via hypertext transfer protocol (http) / file transfer protocol (ftp) or from core facilities or internet databases such as GEO or SRA. Compressed or uncompressed FASTQ (.fastq) and SRA (.sra) files can be used. We elected not to use the prealigned BAM (.bam) files to ensure uniform alignment of samples.

For ChIP-Seq and similar experiments, reads are aligned to the genome with Bowtie [13], quality control analysis is conducted and data are summarized in a table (Fig. 2A). In addition to basic statistics (percentages of mapped/unmapped/non-uniquely mapped reads and average fragment length), Wardrobe displays several other quality control measures. Base frequency plots are used to estimate adapter contamination – a frequent occurrence in low-input ChIP-Seq experiments (Fig. 2B). Average tag density profiles can be used to estimate ChIP enrichment for promoter proximal histone modifications (e.g. histone 3 lysine residue 4 trimethylation [H3K4me3], Fig. 2F). The genome browser can be used to visually compare results to other experiments in the database (Fig. 2C). ChIP-Seq results are displayed on the genome browser as coverage per million reads mapped. For paired-end reads, coverage is calculated as the number of fragments covering each base pair (bp). To obtain coverage for single-read experiments, average fragment length is calculated by MACS [16], and individual reads are extended to this length in the 3' direction. Islands of enrichment identified by MACS are displayed both on the browser (Fig. 2C) and as a table (Fig. 2D) together with the nearest genes. Use of different parameters or pipelines for different antibodies (e.g. "broad peaks" MACS option for H3K27me3) is possible. The distribution of the islands between genomic areas (promoters, exons, etc.) is displayed as a stacked bar graph (Fig. 2E).

For RNA-Seq analysis, reads are aligned with RNA STAR [14] using the appropriate RefSeq transcriptome. The quality control tab displays the number of reads aligned within and outside the transcriptome. The percentage of the reads mappable to ribosomal (r)DNA is displayed to estimate the quality of rRNA depletion (Fig. 2G). Data are deposited on the browser, and RPKMs are calculated for each transcript (algorithm to be described elsewhere) (Fig. 2H,I). Depending on the application, RPKM values can be presented for each transcript or summed up for each TSS (for gene expression studies) or for each gene (for functional studies, e.g. Gene Ontology).

Advanced analysis

If satisfied with the quality of data obtained from sequencing, a user can proceed to advanced analysis, which involves integration of information from multiple experiments. For gene expression analysis, the typical task is identifying differentially expressed genes. We elected to incorporate the DESeq1/2 algorithm [17, 19] for this purpose because it does not require recreating transcript models and does not make many assumptions. In order to perform gene expression profiling, a user can define replicates and utilize the DESeq algorithm to calculate p-values and fold changes. On the basis of DESeq results, lists of genes whose expression changes can be created within Wardrobe using expression levels, fold change, or p/q-values, as well as other parameters, and downloaded, if needed, in a table form for further analysis (e.g., gene set enrichment analysis) (Fig. 3A).

The gene sets can also be used to create average tag density profiles and heatmaps within Wardrobe (Fig. 3B-F). Average tag density profiles are used to compare the enrichment of histone modifications or other proteins around the TSS or the gene bodies between different gene sets. Often gene bodies used to estimate enrichment, for instance when comparing the levels of positive marks, such as H3K4me3 between expressed and silent genes. Heatmaps provide similar information but allow comparisons of modifications between individual genes. Statistical comparison of tag densities between groups of genes can be performed using Mann-Whitney-Wilcoxon (MWW) tests (Fig. 3C and D). All graphs can be downloaded in publication-quality SVG format.

For ChIP-Seq, the task is usually the identification of areas that have different levels of binding between samples. The difficulty here is that the signal-to-background ratio (enrichment) is usually slightly different between ChIP-Seq experiments; thus, several assumptions have to be made in order to compare islands of enrichment. Wardrobe uses the MAnorm algorithm [18], which assumes that modifications do not change in the majority of areas. This allows MAnorm to adjust for differential levels of enrichment between experiments. The lists of islands, fold changes and accompanying p-values are presented in table form, and islands can be viewed in the browser with the push of a button (Fig. 3G).

Epigenetic changes during Th1 differentiation of human T cells

To demonstrate the ability of Wardrobe to integrate data obtained from different sources and using different sequencing technologies, we have performed analysis of gene expression and chromatin environment changes during differentiation of human naïve T cells into T helper type 1 (Th1) cells. For this purpose, we utilized a gene expression and chromatin dataset published by Hawkins et al. (SRA082670)[20]. The study included RNA-Seq data for resting naïve T cells ("T helper precursor cells [Thp]" in the paper) and T cells activated in Th1 conditions for 72 hours in triplicate and ChIP-Seq data for T cells activated in Th1 conditions. We wanted to identify chromatin changes that occur during Th1 differentiation of T cells, but the Hawkins et al. study did not include chromatin profiles for resting naïve cells. For this reason, we used H3K4me3 ChIP-Seq data for resting naïve T cells obtained for a separate project in our laboratory. In this study, RNA-Seq data was obtained using a Helicos platform, whereas ChIP-seq libraries were sequenced on Illumina.

After entering the sample descriptions and links to .sra files from the SRA database into the system, Wardrobe downloaded the dataset and performed basic analysis (Fig. 2A). ChIP-Seq data demonstrated the expected percentage of reads mapped (Fig. 2B), average tag density profiles showed high enrichment at promoters (Fig. 2F), MACS identified ~48 thousand and 79 thousand islands (naïve T cells and Th1 cells, respectively, Fig. 2D), the majority of which (68-77%) were located at promoters (Fig. 2E). However, RNA-Seq results demonstrated poor mapping to the human transcriptome, especially for naïve T cell samples (2-6%) (Fig. 2G), whereas 35% of reads were mapped outside the annotated transcriptome. Altogether, these quality control measurements led us to conclude that the RNA-Seq libraries were severely contaminated by genomic DNA and that the RPKM values are likely to be inflated by several RPKMs, in particular in the naïve samples. Nonetheless, we proceeded with the advanced analysis.

We next performed comparison of gene expression using DESeq2. Replicates were defined, genes were grouped by common TSS and differentially expressed genes were identified. DESeq results (Fig. 3A) were used to define lists of genes that were expressed or silent in both Thp and Th1 cells or that were induced during differentiation. These lists were used to create average tag density profiles in both naïve and Th1 cells (Fig. 3B,D,E). As demonstrated in the graphs and from MWW statistical analysis (Fig. 3D), expressed genes have higher levels of H3K4me3 at their promoters than the silent genes. Inducible genes had intermediate levels of this modification in naïve cells, in which these genes were silent, suggesting that H3K4me3 poises inducible genes for expression. The same conclusion can be made upon examining tag density heatmaps (Fig. 3B). Upon induction, the H3K4me3 levels of the genes in the induced list increased to the level of the genes in the expressed list.

The H3K4me3 ChIP-Seq experiments worked well in both naïve and Th1 cells; however, if we compare the height of the TSS peaks of H3K4me3 between naïve and Th1 cells (Fig. 2D and data not shown), we can conclude that the enrichment in the naïve experiment is slightly higher. This difference in enrichment would affect the peak calling, meaning that the peaks cannot be compared directly between the two cell types. To overcome this obstacle, Wardrobe uses the MAnorm algorithm, which compares the read numbers between islands while adjusting for enrichment [18]. MAnorm produces fold changes and p-values for the peaks, and the areas of interest can be viewed on the browser (Fig. 3F).

Implementation

Wardrobe is accessible via Google Chrome, Safari and Firefox browsers. The user interface is web based and utilizes HTML5 and JavaScript technologies. To speed up the development process, EXTJS and D3 frameworks were used. On the server, Apache with PHP is used to process user's requests. Linux or MacOSX native job schedulers are used to run Python pipelines. For stability, all pipelines have separate queues and process statuses. Pipeline output is stored in the SQL database with the exception of BAM files. These precomputed data are accessible by third-party software like RStudio that allows analysis that is not included in Wardrobe. There are no specific hardware limitations for Wardrobe. We have installed it on both a Linux server and Mac Pro desktop and laptop computers. An average Intel Core i7 computer with 32 gigabytes of RAM and a SATA HDD; more than 100 M read/write speed) will analyze one ChIP-Seq or RNA-Seq experiment within 2 hours.

Current code and setup instructions are available at https://code.google.com/p/genome-tools/. A limited-functionality demo version is available at http://demo.wardrobe.porter.st/ems .

In summary, we have developed an automated system for storage and analysis of NGS data. The Wardrobe system can be easily installed on Mac or Linux computers and provides data analysis solutions for the whole laboratory or institution.

List of abbreviations used

ATAC-Seq	assay for transposase-accessible chromatin sequencing
ChIP	chromatin immunoprecipitation
ChIP-Seq	chromatin immunoprecipitation sequencing
D3	data-driven documents
DNase-Seq	DNase I hypersensitive sites sequencing
Ext JS	extension JavaScript
ftp	file transfer protocol
GEO	Gene Expression Omnibus
GRO-Seq	global run-on sequencing
HIPAA	Health Insurance Portability and Accountability Act of 1996
HTML	hypertext markup language
MACS	model-based analysis of ChIP-Seq
MAnorm	a model for quantitative comparison of ChIP-Seq datasets
Me3	trimethylation
MNase-Seq	micrococcal nuclease sequencing
MWW	Mann-Whitney-Wilcoxon
NGS	next-generation sequencing

PHP	PHP: hypertext preprocessor
RAM	random-access memory
RefSeq	National Center for Biotechnology Information Reference Sequence
RPKMs	reads per kilobase of transcript per million reads mapped
RNA-Seq	ribonucleic acid sequencing
SATA HDD	serial ATA hard disk drive
SRA	Sequence Read Archive (formerly known as Short Read Archive)
STAR	spliced transcripts alignment to a reference
SVG	scalable vector graphics
TSS	transcription start site(s)
Th1	T helper type 1
Thp	T helper precursor cells
UCSC	University of California, Santa Cruz

Competing interests

The authors declare no competing financial interests.

Authors' contributions

AK and AB designed the system, AK wrote the program, AK and AB wrote the paper.

Authors' information

Andrey V. Kartashov, M.S.[1] and Artem Barski, Ph.D. [1,2]

[1]Division of Allergy and Immunology, [2]Division of Human Genetics, Cincinnati Children's Hospital Medical Center and Department of Pediatrics, College of Medicine, University of Cincinnati, Cincinnati, OH

Acknowledgements

Authors would like to thank Shawna Hottinger (CCHMC) for editorial assistance.

The work was supported in part by the Children's Hospital Research Foundation and NHLBI NIH career transition award to A.B. (K22 HL098691).

References

1. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-Seq**. *Nat Methods* 2008, **5**:621–628.

2. Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K: **High-resolution profiling of histone methylations in the human genome.** *Cell* 2007, **129**:823–37.

3. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE: **High-resolution mapping and characterization of open chromatin across the genome**. *Cell* 2008, **132**:311–322.

4. Schones DE, Cui K, Cuddapah S, Roh T-Y, Barski A, Wang Z, Wei G, Zhao K: **Dynamic regulation of nucleosome positioning in the human genome.** *Cell* 2008, **132**:887–98.

5. Core LJ, Waterfall JJ, Lis JT: **Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters.** *Science* 2008, **322**:1845–8.

6. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ: **Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position.** *Nat Methods* 2013, **10**:1213–8.

7. **Genespring NGS** [http://www.genomics.agilent.com/en/NGS-Analysis-Software/GeneSpring-NGS-versions-12-5-12-6-1-/?cid=cat170010&tabId=AG-PR-1062]

8. **Partek** [www.partek.com]

9. **Golden Helix** [http://www.goldenhelix.com]

10. **Illumina basespace** [https://basespace.illumina.com/home/index]

11. Goecks J, Nekrutenko A, Taylor J: **Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences.** *Genome Biol* 2010, **11**:R86.

12. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D: **The human genome browser at UCSC.** *Genome Res* 2002, **12**:996–1006.

13. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome Biol* 2009, **10**:R25.

14. Dobin A, Davis C a, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR: **STAR: ultrafast universal RNA-seq aligner.** *Bioinformatics* 2013, **29**:15–21.

15. **FASTX** [http://hannonlab.cshl.edu/fastx_toolkit/index.html]

16. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS: **Model-based analysis of ChIP-Seq (MACS).** *Genome Biol* 2008, **9**:R137.

17. Anders S, Huber W: **Differential expression analysis for sequence count data.** *Genome Biol* 2010, **11**:R106.

18. Shao Z, Zhang Y, Yuan G-C, Orkin SH, Waxman DJ: **MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets.** *Genome Biol* 2012, **13**:R16.

19. Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, Robinson MD: **Count-based differential expression analysis of RNA sequencing data using R and Bioconductor.** *Nat Protoc* 2013, **8**:1765–86.

20. Hawkins RD, Larjo A, Tripathi SK, Wagner U, Luu Y, Lönnberg T, Raghav SK, Lee LK, Lund R, Ren B, Lähdesmäki H, Lahesmaa R: **Global chromatin state analysis reveals lineage-specific enhancers during the initiation of human T helper 1 and T helper 2 cell polarization.** *Immunity* 2013, **38**:1271–84.

Figure legends

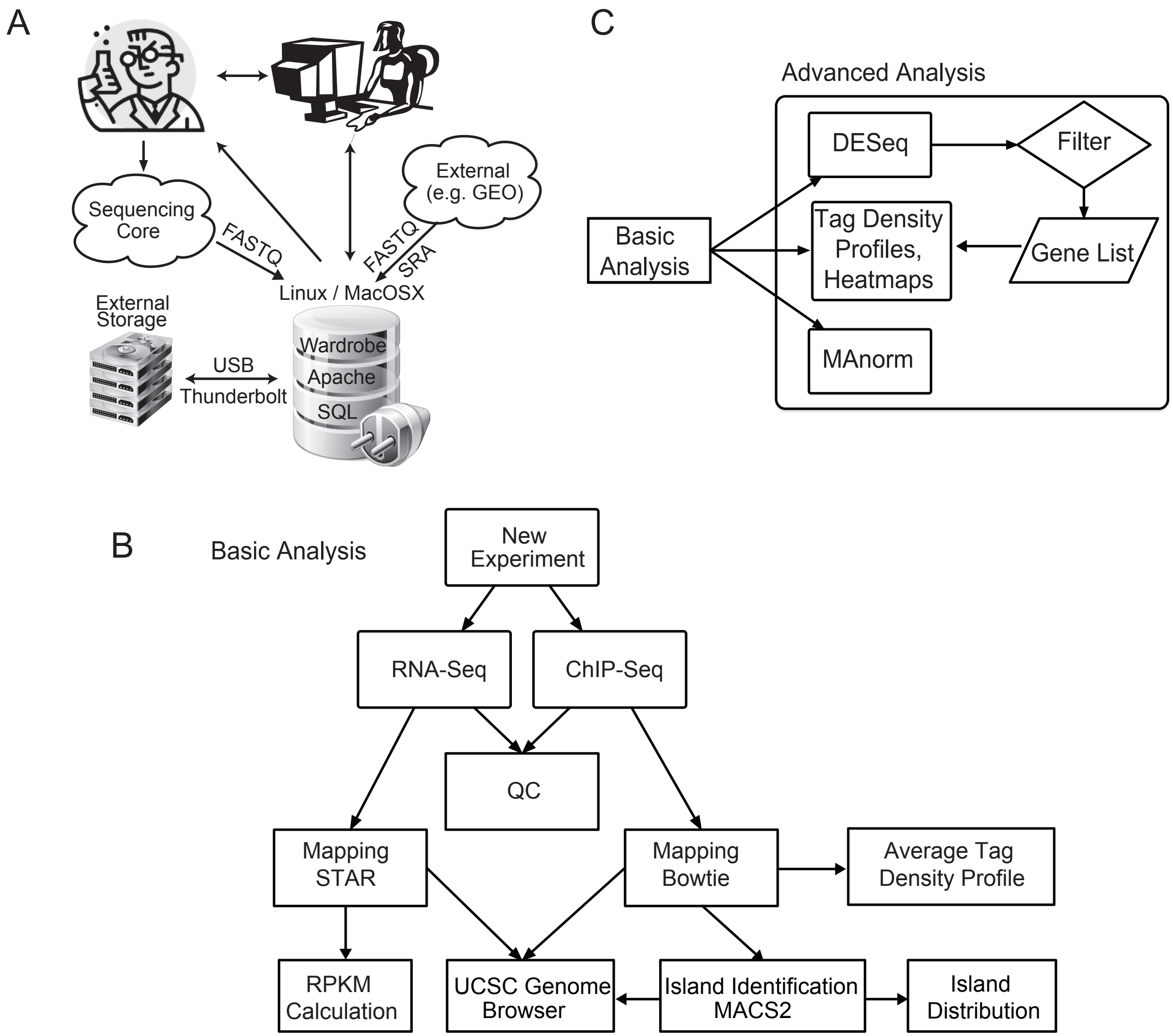
Figure 1. Wardrobe overview and pipelines. A. The Wardrobe server can be set up on a Linux or Mac computer attached to a consumer-class storage array, typically within a local institutional network. Biologists can use Wardrobe to upload data from a sequencing core or a public database and promptly receive quality control data, view the results in the browser and perform some of the analysis without the assistance of bioinformaticians. Bioinformaticians can access the precomputed data in Wardrobe's SQL database to perform further analysis. **B.** Basic analysis pipelines. The flow diagram shows the tools used in the basic analysis pipelines for RNA-Seq and ChIP-Seq data. **C.** Advanced analysis allows the user to identify differentially expressed genes using DESeq, create gene lists and use these list to generate average tag density profiles and heatmaps. Differentially bound areas can be identified with MAnorm.

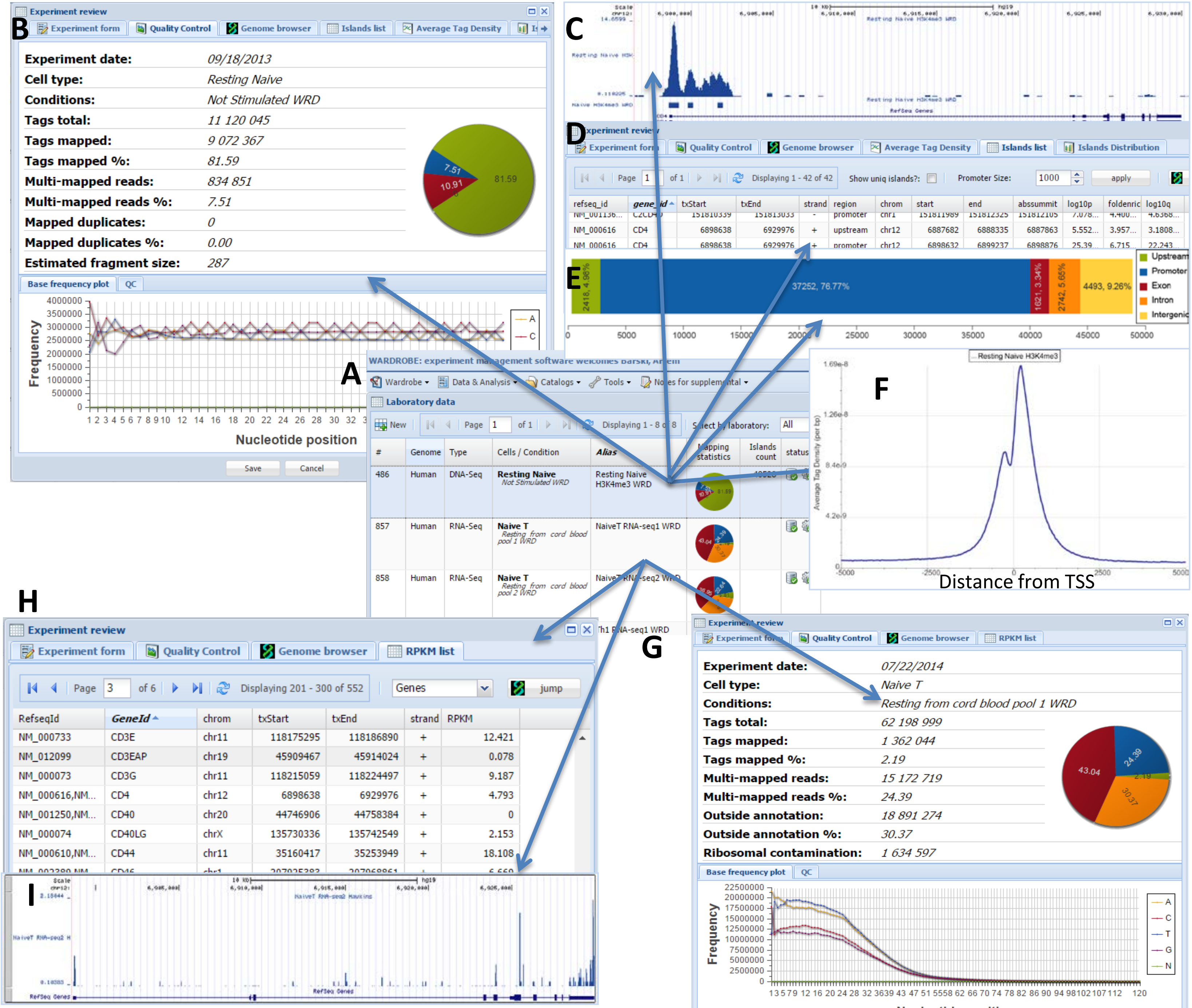
Figure 2. Basic analysis. A. The laboratory data table shows the list of experiments available to the user. **B.** Shown is the quality control tab for a ChIP-Seq experiment; the spiky base frequency plot indicates minor adapter contamination. **C.** Shown is the ChIP-Seq browser shot for the *CD4* gene, visible on the genome browser tab. **D.** The island list tab shows locations of islands and the nearest genes. **E.** The island distribution bar graph, visible from the islands distribution tab, shows that the majority of H3K4me3 islands are located at promoters. **F.** The average tag density profile, visible from the average tag density tab, shows enrichment around the TSS. **G.** Shown is the quality control tab for an RNA-Seq experiment. The base frequency plot shows an AT bias, suggesting DNA contamination and read-length variation characteristic of Helicos sequencing. **H.** Shown is the RPKM table available from the RPKM list tab for an RNA-Seq experiment. **I.** Shown is the RNA-Seq browser for the *CD4* gene, visible from the genome browser tab.

Figure 3. Advanced analysis. A. The table displays the results of DESeq2 differential gene expression analysis, showing induction of TNF upon Th1 differentiation. **B.** The heatmaps show expression and

H3K4me3 tag density in the three examined sets of genes (Silent, Expressed, Induced) in naïve T cells and Th1-differentiated T cells. **C,E,F** Shown are the H3K4me3 average tag density profiles for the gene body (**C**) or around the TSS (**E,F**) in naïve T cells and Th1-differentiated T cells. **D.** The box plot shows the distribution of H3K4me3 tag densities for the three gene sets within the area shaded in panel C (Silent, Expressed, Induced). MWW p-values are shown below the box plot. **G.** Shown is identification of differentially modified areas using MAnorm. Notice the appearance of modifications ("description" column) at Th1 characteristic cytokine IFNG.

Kartashov Figure 1





B

Experiment date:	*09/18/2013*
Cell type:	*Resting Naive*
Conditions:	*Not Stimulated WRD*
Tags total:	*11 120 045*
Tags mapped:	*9 072 367*
Tags mapped %:	*81.59*
Multi-mapped reads:	*834 851*
Multi-mapped reads %:	*7.51*
Mapped duplicates:	*0*
Mapped duplicates %:	*0.00*
Estimated fragment size:	*287*

D

Page 1 of 1 Displaying 1 - 42 of 42 Show uniq islands?: □ Promoter Size: 1000 apply

refseq_id	gene_id	txStart	txEnd	strand	region	chrom	start	end	abssummit	log10p	foldenric	log10q
NM_001136...	C2CD4D	151810339	151813033	-	promoter	chr1	151811989	151812325	151812105	7.078...	4.400...	4.6368...
NM_000616	CD4	6898638	6929976	+	upstream	chr12	6887682	6888335	6887863	5.552...	3.957...	3.1808...
NM_000616	CD4	6898638	6929976	+	promoter	chr12	6898632	6899237	6898876	25.39...	6.715...	22.243...

A

Page 1 of 1 Displaying 1 - 8 of 8 Select by laboratory: All

#	Genome	Type	Cells / Condition	*Alias*	Mapping statistics	Islands count	status
486	Human	DNA-Seq	**Resting Naive** *Not Stimulated WRD*	Resting Naive H3K4me3 WRD		40520	
857	Human	RNA-Seq	**Naive T** *Resting from cord blood pool 1 WRD*	NaiveT RNA-seq1 WRD			
858	Human	RNA-Seq	**Naive T** *Resting from cord blood pool 2 WRD*	NaiveT RNA-seq2 WRD			
				Th1 RNA-seq1 WRD			

G

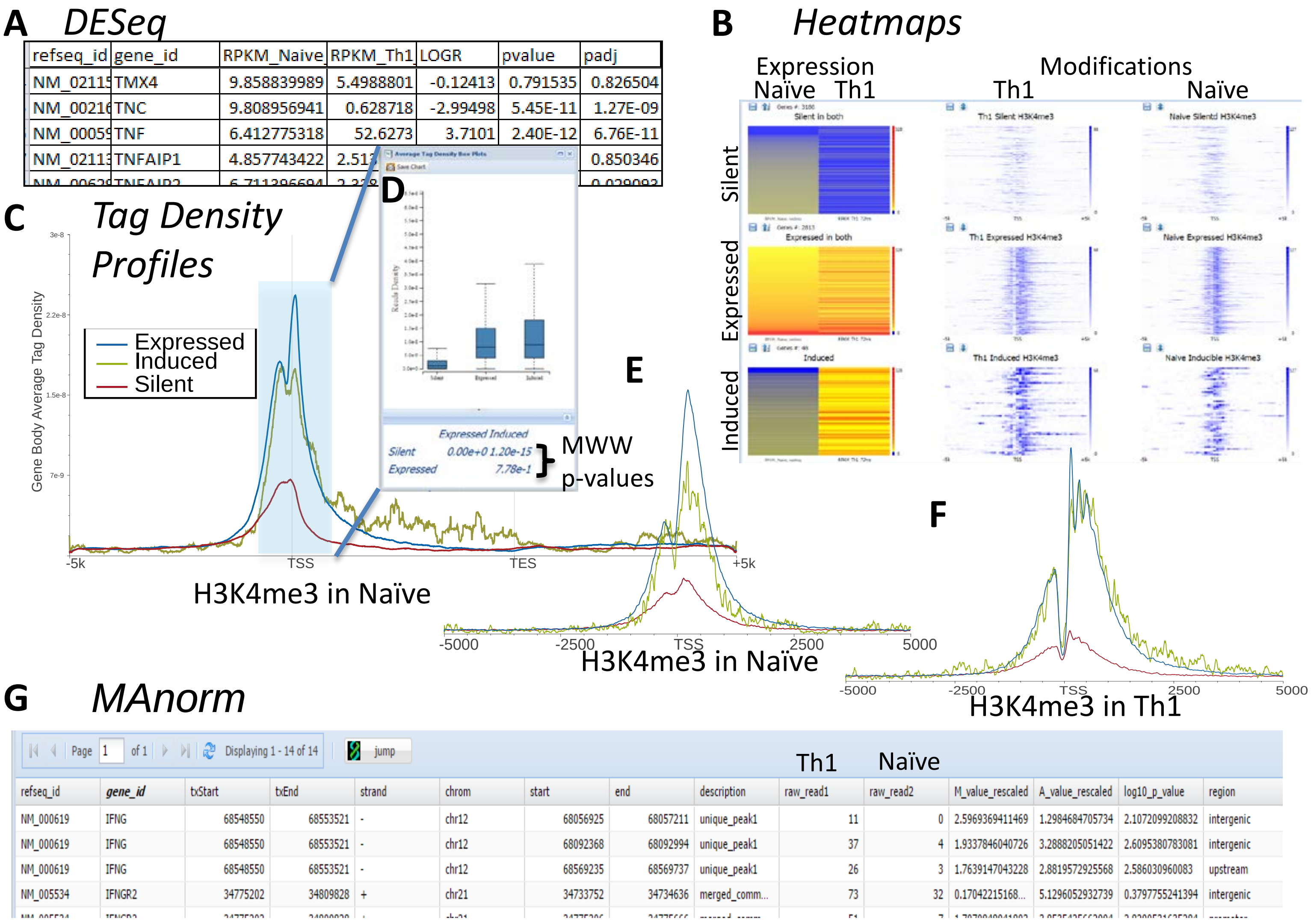
Experiment date:	*07/22/2014*
Cell type:	*Naive T*
Conditions:	*Resting from cord blood pool 1 WRD*
Tags total:	*62 198 999*
Tags mapped:	*1 362 044*
Tags mapped %:	*2.19*
Multi-mapped reads:	*15 172 719*
Multi-mapped reads %:	*24.39*
Outside annotation:	*18 891 274*
Outside annotation %:	*30.37*
Ribosomal contamination:	*1 634 597*

H

Page 3 of 6 Displaying 201 - 300 of 552 Genes

RefseqId	GeneId	chrom	txStart	txEnd	strand	RPKM
NM_000733	CD3E	chr11	118175295	118186890	+	12.421
NM_012099	CD3EAP	chr19	45909467	45914024	+	0.078
NM_000073	CD3G	chr11	118215059	118224497	+	9.187
NM_000616,NM...	CD4	chr12	6898638	6929976	+	4.793
NM_001250,NM...	CD40	chr20	44746906	44758384	+	0
NM_000074	CD40LG	chrX	135730336	135742549	+	2.153
NM_000610,NM...	CD44	chr11	35160417	35253949	+	18.108
NM_002389,NM...	CD46	chr1	207925383	207968861	+	6.669

Kartashov Figure 2



refseq_id	gene_id	RPKM_Naive	RPKM_Th1	LOGR	pvalue	padj
NM_02115	TMX4	9.858839989	5.4988801	-0.12413	0.791535	0.826504
NM_00216	TNC	9.808956941	0.628718	-2.99498	5.45E-11	1.27E-09
NM_00059	TNF	6.412775318	52.6273	3.7101	2.40E-12	6.76E-11
NM_02113	TNFAIP1	4.857743422	2.513			0.850346

Page 1 of 1 Displaying 1 - 14 of 14 jump

refseq_id	gene_id	txStart	txEnd	strand	chrom	start	end	description	raw_read1	raw_read2	M_value_rescaled	A_value_rescaled	log10_p_value	region
NM_000619	IFNG	68548550	68553521	-	chr12	68056925	68057211	unique_peak1	11	0	2.5969369411469	1.2984684705734	2.1072099208832	intergenic
NM_000619	IFNG	68548550	68553521	-	chr12	68092368	68092994	unique_peak1	37	4	1.9337846040726	3.2888205051422	2.6095380783081	intergenic
NM_000619	IFNG	68548550	68553521	-	chr12	68569235	68569737	unique_peak1	26	3	1.7639147043228	2.8819572925568	2.586030960083	upstream
NM_005534	IFNGR2	34775202	34809828	+	chr21	34733752	34734636	merged_comm...	73	32	0.17042215168...	5.1296052932739	0.3797755241394	intergenic

Kartashov Figure 3