Comparative analysis in bat and mouse identifies wing-specific genes and regulatory regions

Guy Kelman, Walter L. Ecklabar, Nicola Illing, Nadav Ahituv, Tommy Kaplan

Abstract

The bat is the only mammal capable of powered flight. To identify the molecular mechanisms underlying the development of bat forelimbs into wings, we studied comparative gene expression and chromatin data from mouse and bat.

We have collected gene expression (mRNA-seq) and genome-wide regulatory chromatin data (ChIP-seq for the active H3K27ac, and the repressive H3K27me3 marks) from developing forelimbs and hindlimbs, during three early developmental stages of the Natal long-fingered bat (*Miniopterus natalensis*) and the mouse (*Mus musculus*).

We then developed a computational model to compare expression or ChIP data across organism, limbs and time. We identified a set of ~2000 genes and ~1500 genomic loci that present significant forelimb-specific patterns in the bat.

As we show, our analysis reveals with great confidence genes and DNA regions that are well known for their roles during limb development, as well as many less studied ones.

This study offers interpretation of how local changes in regulatory elements are translated into dramatic functional and phenotypic differences. Furthermore, the signalling pathways involved in limb development are well conserved among species, suggesting that identification of novel genes and distal regulatory regions could shed light on limb malformations in humans.

The developing bat wing

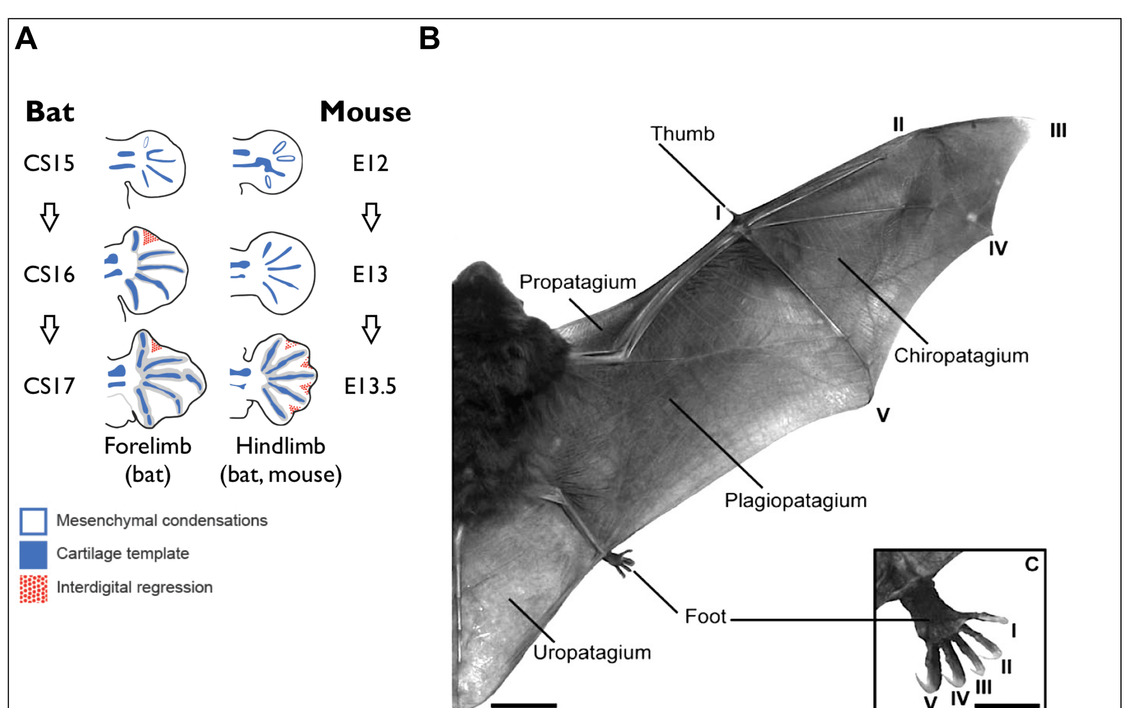
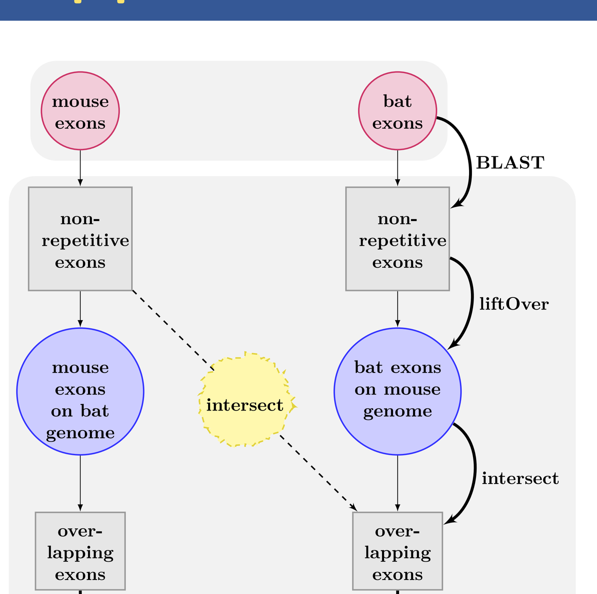
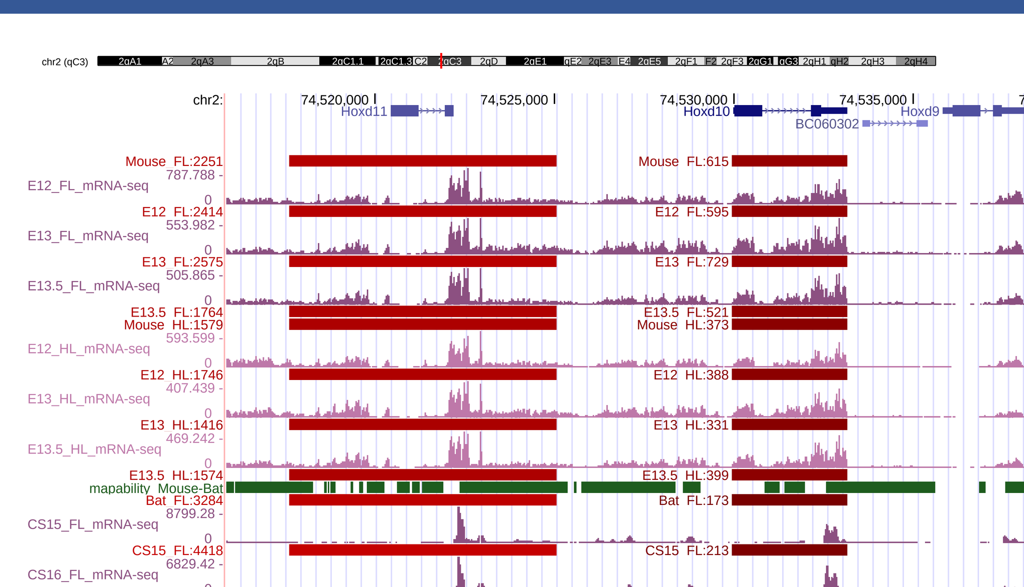


Figure 1. Limb development in bat and mouse. (A) Forelimb development in bats (*Miniopterus natalensis*), from developmental stage CS15 (top) to CS17 (bottom). Shown are mesenchymal condensations (white), cartilages (blue) and tissue interdigital regression (red), which are combined to form the wing. In parallel, we present hindlimb development in bat and mouse (developmental stages E12, E13 and E13.5). Adapted from (Eckalbar *et al.*, manuscript under review) **(B)** Forelimb and hindlimb anatomy. Shown is ventral view of adult bat (*M. natalensis*) and its wing, compared to hindlimb (inset C), with thumb (I) and digits (II-IV) marked (Adapted from Hockman *et al.* [1])

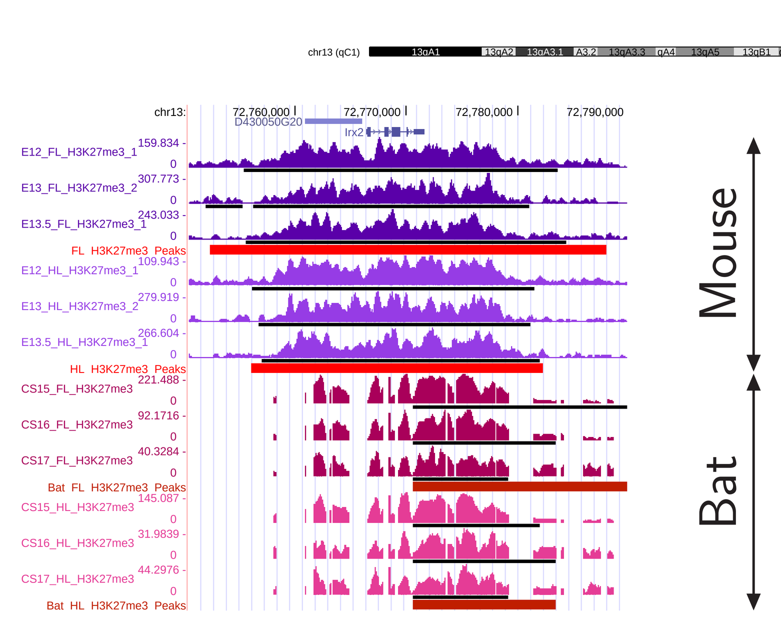
The pipeline



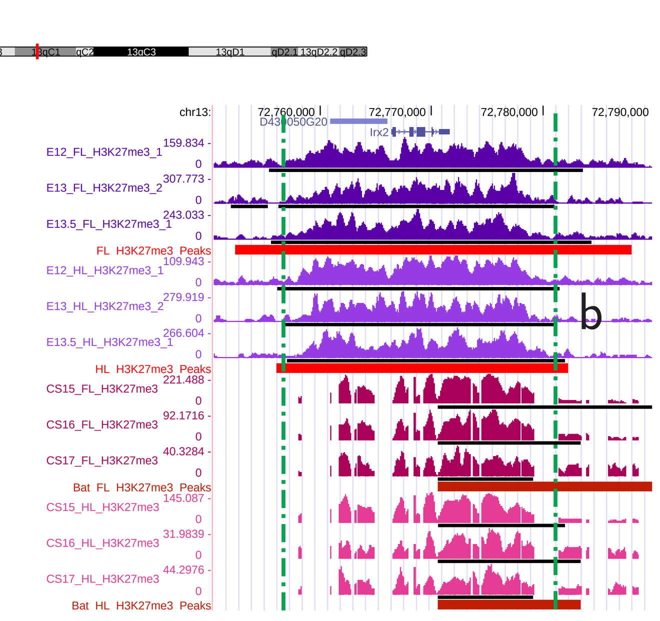
transcriptome profiling and differential expression



mouse and bat RNA-Seq tracks in the HoxD cluster



mouse and bat ChIP H3K27me3 tracks



intersect from a basepoint **b** at E13 hindlimb

The pipeline demonstrates extraction and differential analysis of RNA-Seq data. Similarly, peaks from genome-wide ChIP data will come in place of the exon assay (red circles).

Analysis of variance

We use a statistical design to predict the read-counts or peak heights that were observed in any specific expression or chromatin activation level. Our null hypothesis is that as first order approximation, bat and mouse are indistinguishable.

H_0 the zero-model (reduced model)

Main effects: *Species* (Bat vs. Mouse), *Stage* (E12.0, E13.0, E13.5) and *Tissue* (forelimb vs. hindlimb).

Expect: regions where forelimb is differentiated in activation levels from hindlimb in bat respective to the activation levels in mouse.

```
design:  Tissue + Species + Stage
```

H_1 the full model

Add interactions: *Stage:Species*, *Tissue:Species*

Expect: greater confidence in the predicted fold change.

```
design:  Tissue + Species + Stage + Stage:Species + Tissue:Species
```

To resolve a differential model we test the likelihood ratio between H_0 and H_1, which would reveal differential patterns due to the interaction terms.

Results

RNA-Seq

RNA-Seq data uncovers 1537 wing-specific (differential) genes out of 12745 ranges that were successfully mapped between the animals.

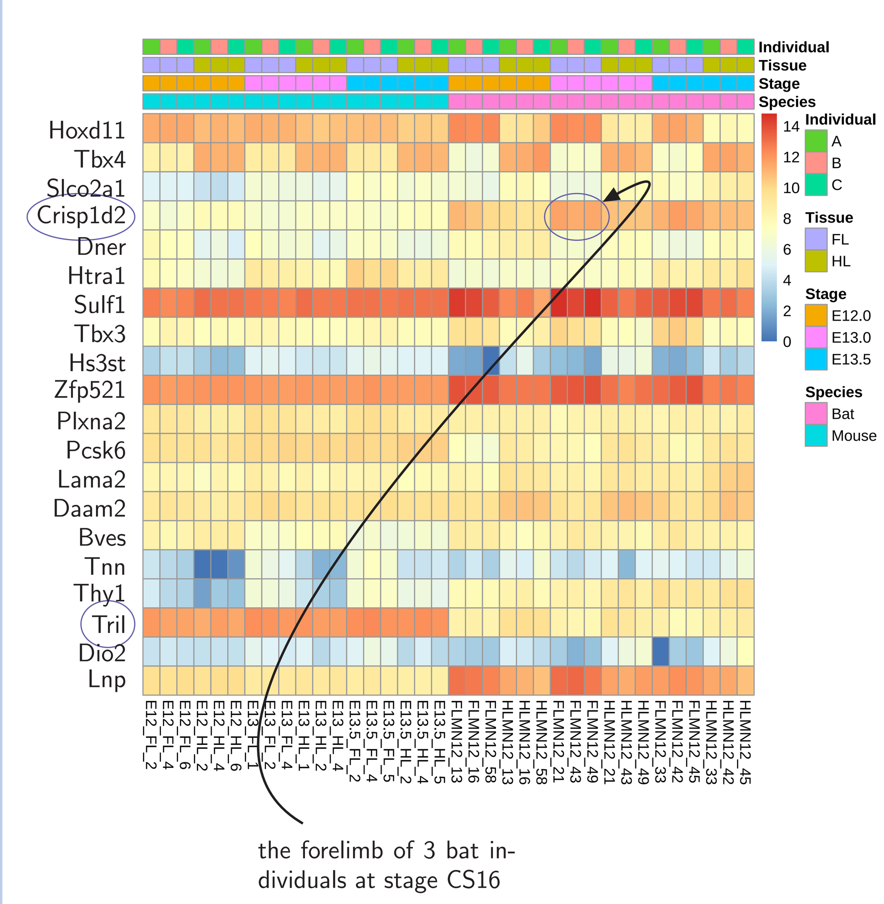
ChIP-Seq

base-point	H3K27ac			H3K27me3		
	int	found	free	int	found	free
CS15FL	8045	5499	6574	4595	2461	5280
CS15HL	7435	5078	6097	4501	2383	6096
CS16FL	7263	4968	5791	2947	1707	2751
CS16HL	5553	3799	3209	2349	1321	1030
CS17FL	6503	4420	4861	3073	2099	1166
CS17HL	8313	5506	7448	3469	2342	1646

base-point	H3K27ac			H3K27me3		
	int	found	free	int	found	free
E12.0FL	3553	2759	1506	1981	1220	1540
E12.0HL	5887	4775	3768	2399	1409	2348
E13.0FL	3525	2621	2570	2823	1543	2632
E13.0HL	5575	3927	4480	6899	3964	10686
E13.5FL	5119	3800	7944	1445	857	1330
E13.5HL	5019	3779	6646	1579	985	1740

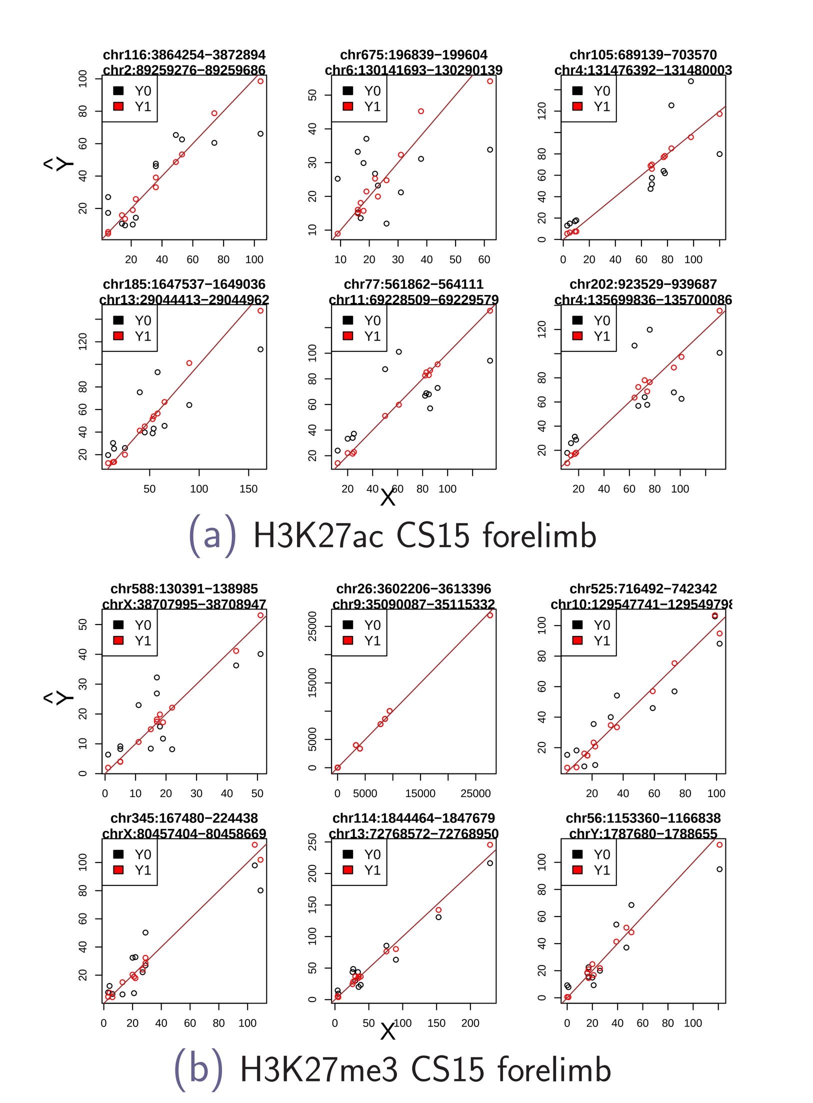
A tabular view of **ChIP** data. In each basepoint, we located **int** intersected regions, of which the result of the model's analysis is given in **found**. The **free** giveaways are peaks that did not intersect in the other species.

Change in expression pattern



A heat map of the top-20 differentially regulated genes that were discovered by the full model. Side-by-side mouse and bat expression counts visualize various operation patterns of wing-specific genes. **Examples**: the bat exhibits two orders of magnitude greater increase in expression of *Crisp1d2* over the mouse, and prominently in the earlier stage forelimbs. *Tril* expression is strictly lower in the bat, especially in the forelimbs.

Confidence in the predictions



Red (Y1) and black (Y0) dots are 12 predictions ($\hat{Y}$) to the measured peaks (X) in one of 12 conditions. Generally Y1 (full model) gives better predictions than Y0 (reduced model). The vertical distance between a pair of red-black is a proxy for the confidence gain in the full model. Each panel gives a single ChIP peak, which chromosomal coordinates are in the panel titles (top for bat and bottom for mouse).

Discussion

Takeaway: This methodology breaks down in great genomic detail the epigenetics that give rise to wing development.

- Tissue, and time-specific analysis uncovers the set of genes and chromatin regions that operate differentially during limb development causing the bat to grow wings, when other animals don't.
- Following this computational study, our group will enter an experimental phase to check the effect of subsets of these genes and regions.

References

- Betty M Booker et al. "Bat accelerated regions identify a bat forelimb specific enhancer in the HoxD locus". In: *PLoS Genet* 12.3 (2016), e1005738.
- Walter L Eckalbar et al. "Transcriptomic and epigenomic characterization of the developing bat wing". In: *Nature genetics* 48 (Mar. 2016), pp. 528–536.
- Walter Ecklabar and Nadav Ahituv. *NCBI Miniopterus natalensis Annotation Release 100*. NCBI Genome Annotation database. Apr. 2016.
- W James Kent et al. "BigWig and BigBed: enabling browsing of large distributed datasets". In: *Bioinformatics* 26.17 (2010), pp. 2204–2207.
- W James Kent et al. "The human genome browser at UCSC". In: *Genome research* 12.6 (2002), pp. 996–1006.
- Michael Love, Simon Anders, and Wolfgang Huber. "Differential analysis of count data–the DESeq2 package". In: *Genome Biology* 15 (2014), p. 550.

HUJI - UCSF - UCT superk@cs.huji.ac.il