Kamps-Hughes N, Preston JL, Randel MA, Johnson EA. 2015. Genome-wide identification of hypoxia-induced enhancer regions. PeerJ 3:e1527 https://doi.org/10.7717/peerj.1527

Genome-wide identification of hypoxia-induced enhancer regions

Nick Kamps-Hughes, Jessica L Preston, Melissa A Randel, Eric A Johnson

Here we present a genome-wide method for *de novo* identification of enhancer regions and apply it to find enhancers that have increased activity after hypoxia. The method links fragmented genomic DNA to the transcription of randomer molecule identifiers and measures the functional enhancer activity of the library by massively parallel sequencing. We transfected a Drosophila melanogaster library into S2 cells in normoxia and hypoxia, and assayed 4,599,881 genomic DNA fragments in parallel. The locations of the enhancer regions strongly correlate with genes up-regulated after hypoxia and previously described enhancers. Novel enhancer regions were identified and integrated with RNAseq data and transcription factor motifs to describe the hypoxic response on a genome-wide basis as a complex regulatory network involving multiple stress-response pathways. This work provides a novel method for high-throughput assay of enhancer activity and the genome-scale identification of hypoxia-activated enhancers in Drosophila.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1393v1 | CC-BY 4.0 Open Access | rec: 25 Sep 2015, publ: 25 Sep 2015

Genome-wide identification of hypoxia-induced enhancer regions

Nick Kamps-Hughes[1], Jessica L Preston[1], Melissa A Randel[1] and Eric A Johnson[1]
[1]Institute of Molecular Biology, University of Oregon, Eugene, Oregon, 97403, USA

Corresponding author:
Nick Kamps-Hughes
590 Merritt Ave #32, Oakland, CA, 94610, USA
Email address: nkampshughes@gmail.com

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1393v1 | CC-BY 4.0 Open Access | rec: 25 Sep 2015, publ: 25 Sep 2015

Genome-wide identification of hypoxia-induced enhancer regions

Nick Kamps-Hughes[1], Jessica L Preston[1], Melissa A Randel[1] and Eric A Johnson[1]
[1]Institute of Molecular Biology, University of Oregon, Eugene, Oregon, 97403, USA

ABSTRACT

Here we present a genome-wide method for *de novo* identification of enhancer regions and apply it to find enhancers that have increased activity after hypoxia. The method links fragmented genomic DNA to the transcription of randomer molecule identifiers and measures the functional enhancer activity of the library by massively parallel sequencing. We transfected a Drosophila melanogaster library into S2 cells (contributed by Ken Prehoda lab, University of Oregon) in normoxia and hypoxia, and assayed 4,599,881 genomic DNA fragments in parallel. The locations of the enhancer regions strongly correlate with genes up-regulated after hypoxia and previously described enhancers. Novel enhancer regions were identified and integrated with RNAseq data and transcription factor motifs to describe the hypoxic response on a genome-wide basis as a complex regulatory network involving multiple stress-response pathways. This work provides a novel method for high-throughput assay of enhancer activity and the genome-scale identification of hypoxia-activated enhancers in Drosophila.

INTRODUCTION

Gene expression is differently regulated in different cell types and in response to changes to environmental conditions. This regulation is achieved in part by the activity of transcriptional enhancers[1-5], specific DNA sequences that bind transcription factors to control the rate of transcription initiated at nearby promoters. Even for relatively simple processes, such as the acute response to changes in oxygen availability, the identification and characterization of the enhancers used to shift the network of gene expression to a new mode remains limited.

The transcription factor hypoxia-inducible factor-1 (HIF-1) is directly inhibited by the presence of cellular oxygen via protein degradation of the HIF-1α subunit[6]. Once stabilized, HIF-1α moves to the nucleus and up-regulates the transcription of target genes. Although HIF-1 remains a central regulator in models of how cells respond after experiencing low oxygen[7,8], more

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1393v1 | CC-BY 4.0 Open Access | rec: 25 Sep 2015, publ: 25 Sep 2015

recently other transcription factors have been implicated in the hypoxic response in a complex network of regulatory events. For example, the immunity response transcription factor NF-KB is also activated by hypoxia and regulates the transcription of HIF-1[9,10], while HIF-1 appears to play a reciprocal role in the regulation of NF-kB targets[11]. Likewise, HIF-1 sensitizes the heat shock response by directly regulating heat shock factor (HSF) transcription during hypoxia. Thus, the broader picture that has emerged is that the stress response transcription factor pathways are not isolated regulatory units but rather cooperate and co-opt each other to modify the cell's functions in a complex manner.

High-throughput sequencing tools have become widespread in gene expression studies[12-14]. For example, RNAseq has become a powerful tool for analyzing differential gene expression by quantifying the RNA abundance of the transcriptome. However, RNAseq does not provide empirical information about the regulatory events leading to a change in transcript abundance. ChIPseq provides information about where transcription factors bind to the genome, but binding events do not always result in an active enhancer or change in the rate of transcription. Another sequencing strategy assays open chromatin conformations[15] as a reliable proxy for enhancers. However, until recently the typical functional assay for enhancers was to clone the putative regulator upstream of a reporter gene driven by a minimal promoter.

Several next-generation sequencing-based methods have been used to dissect the function of individual nucleotides within previously known enhancers[16-19] as well as scan genomic sequence for enhancer activity[20]. These methods have either used UTR tags to assay from thousands to hundreds of thousands of fragments in parallel[16-19] or have had to confine the potential enhancer itself to the UTR in order to assay genome-scale complexities[20]. Here we use a novel variation on these high-throughput enhancer screening methods to identify regions of the Drosophila genome with increased activity under hypoxia. Our technique combines randomly sheared genomic fragments to be assayed for activity with a UTR randomer tag system for highly multiplexed tracking of transcriptional activity. The construct library is modularly synthesized in vitro making the relative placement of construct elements easily mutable. This is in contrast to a similar method called STARR-Seq[20] that requires the potential enhancer itself be placed downstream of the transcription start site. Although enhancers are known to function at variable distance and orientation with respect to a target promoter[21,22] their strength has been shown to be modulated by their position relative to the target promoter[23] and transcriptional read-through has been shown to attenuate their activity[24]. The method in this paper allows the regulatory element to be placed at the discretion of the experimenter. Additionally, the previously published library construction methods[16-20] require microbial propagation of DNA libraries whereas we present a simpler entirely in vitro strategy. The work presented here is the first implementation of a massively parallel reporter assay to study cis-regulatory activity during an environmental stress response. A library of 4,599,881 random 400-500 bp fragments spanning the Drosophila melanogaster genome was used to identify 31 hypoxic enhancer regions. The regions coincide with genes up-regulated under hypoxia and with binding site motifs from multiple transcription factors involved in the hypoxic response. This work provides mechanistic details of the hypoxic response by empirically identifying regulatory regions that drive hypoxic transcription, linking them to target genes from RNAseq differential expression data, and identifying trans-acting factors in silico. This genome-wide scan demonstrates the complexity of the hypoxic response, which involves multiple regulators acting in concert to control the expression of a wide variety of targets.

MATERIALS AND METHODS

All DNA sequencing was performed on the Illumina HiSeq. All PCR reactions contained a final concentration of 400nM of each primer and used Phusion Polymerase in 1X HF buffer. All oligonucleotide sequences are listed in Supplementary File S1.

Library synthesis

The linear reporter library used to assay enhancer activity was constructed entirely *in vitro* (Figure 1A). The sequence space being assayed for enhancer activity, in this case the *Drosophila melanogaster* genome, was sonically sheared to generate random enhancer-sized fragments. Adapter ligation and 5’ PCR addition were used to add the Illumina first-end sequence upstream of the sheared DNA and part of the minimal promoter downstream. 5’ PCR additions are used to add minimal promoter elements, an intron to stabilize mRNAs[25], the 20N randomer tag, and

Illumina paired-end sequence upstream of an arbitrary ORF, in this case GFP. The synthetic minimal promoter used was designed to contain several core motifs and has been shown to function with a wide range of enhancers[26]. The two fragments are then ligated together to create the final construct library pictured in Figure 1A. The reporter library was diluted to a target of 10,000,000 molecules and regenerated by PCR so that the library could be adequately characterized by paired-end sequencing. An aliquot of the reporter library is used for paired-end sequencing to match randomer tags located in the 5' UTR to the non-transcribed genomic region driving their expression. The library is then transfected into cells for massively parallel enhancer assay (Figure 1B).

Drosophila melanogaster strain Oregon-R genomic DNA was sonically sheared using the BioRuptor. 400-500bp fragments were isolated by gel electrophoresis then end-repaired using Blunt Enzyme mix (NEB) and 3' adenylated using Klenow exo- (NEB). This sample was then ligated to an asymmetric adapter with T-overhang composed of annealed oligonucleotides Genomic-Adapter-1 and Genomic-Adapter-2. The ligation product was gel-purified and used as PCR template with primers Illumina P5 and Genomic-R to create a library of molecules containing a random 400-500 bp stretch of *Drosophila melanogaster* genomic sequence between the Illumina end one sequence and the beginning of a synthetic promoter. Separately, The GFP coding sequence followed by the SV40 terminator was PCR amplified from plasmid pGreen-H-Pelican with primers GFP-F and SV40-R. This product was then used as template for a PCR reaction using primers SV40-R and Marker-1-F. This product was then used as template for a PCR reaction using primers SV40-R and Marker-2-F. This product was then used as template for a PCR reaction using primers SV40-R and Marker-3-F to create a library of molecules containing a GFP sequence downstream of a minimal promoter with randomer tag and Illumina paired-end sequences. The genomic sequence-containing library and minimal promoter library were then 3' adenylated and 3' thymidylated respectively with Klenow exo- then ligated together. The heterodimer (1819-1919 bp) was gel-purified and subsequently selected for proper orientation by PCR with primers SV40-R and Illumina P5. To reduce library complexity to a scale that was tractable by paired-end sequencing, DNA was quantified using the Qubit system (Invitrogen) and serially diluted to produce an estimated 10,000,000 molecules that were used as template to regenerate the library by PCR with primers SV40-R and Illumina P5. An aliquot of this library was used as template for a PCR reaction with primers Illumina-P7 and Illumina-P5 to generate a paired-end Illumina-sequencing library such that the first-end sequence contained the beginning of the genomic region and the paired-end sequence contained the corresponding randomer tag (Figure 1A). Aliquots were also used to generate transfectable quantities of the full-length reporter library by PCR amplification of the entire fragment using primers SV40-R and Illumina-P5. The final construct library sequence is available in the supplementary material (Supplementary File S2).

Transfection, RNA extraction, and randomer tag sequencing

Six 5mL flasks were plated to 80% confluency with S2 cells and transfected with Fugene HD and 2.6ug reporter library DNA at a 3:1 ratio. The following day three plates were placed under hypoxia (99.5% N2 and 0.5% O2) for five hours and thirty minutes and three were left in atmospheric conditions. Total RNA from both conditions was extracted using Trizol and treated with DNAse Turbo (Ambion). RNA was converted to cDNA with SuperScipt III first strand synthesis kit (Invitrogen) using oligo dT20 primers. cDNA was used as template for PCR with primers flanking the randomer tag to create an amplicon ready for Illumina sequencing. All PCR reactions used Illumina-P7 reverse primer and the following barcoded forward primers to allow multiplexing: RNA-BC-1 for hypoxic sample 1, RNA-BC-2 for hypoxic sample 2, RNA-BC-3 for hypoxic sample 3, RNA-BC-4 for normoxic sample 1, RNA-BC-5 for normoxic sample 5, RNA-BC-6 for normoxic sample 6. The resulting 178-bp amplicons were combined and sequenced on the Illumina Hiseq.

RNAseq

RNA from the same experiments used to quantify enhancer activity was used for RNAseq. mRNA was purified using Dynabeads (Invitrogen) from 10ug of total RNA and chemically fragmented using Ambion Fragmentation Reagent. cDNA libraries were made with SuperScipt III first strand synthesis kit using random hexamer primers followed by second-strand synthesis with DNA Pol I (NEB). The double stranded DNA was end-repaired using NEB Quick Blunting Kit and 3’ adenylated using Klenow exo-. The samples were ligated to divergent Illumina adapters with in-line barcodes (Hypoxic GGTTC, Normoxic CTTCC) and PCR amplified with Illumina primers. 300-450 bp fragments were gel-purified and sequenced on the Illumina HiSeq (hypoxic condition: Accession SRX467593, normoxic condition: Accession SRX467591). 6,855,528 reads from each sample were aligned to the *Drosophila melanogaster* transcriptome (Flybase, r5.22) using TopHat[27]. The bam outputs were analyzed by cufflinks and the resulting transcripts.gtf files were compared using cuffdiff to identify differentially expressed genes (Supplementary File S3). Some ncRNAs were also analyzed for differential expression. As they are not present in the transcriptome build, RNAseq reads were aligned to each ncRNA using Bowtie2[28] and their expression level is reported by normalized number of aligned reads in each condition.

Computational enhancer activity analysis pipeline

All scripts and a tutorial are available in Supplementary File S4. Paired-end fastq files (Accession SRX468157) linking genomic regions in the first-end read to randomer tags in the paired-end read were parsed to a fasta file with the randomer tag as the sequence name and the genomic sequence as the sequence. This file containing 32,061,029 sequences was aligned to the *Drosophila melanogaster* genome (dm3) using Bowtie2[28]. Reads were processed into a match-list linking randomer tags to the genomic coordinates of their corresponding test sequence.

Randomer tags from hypoxic and normoxic RNA amplicon sequencing were extracted from fastq files (Accessions SRX468694, SRX468097) and experimental replicates were separated by barcode. 18,261,667 randomer tags from hypoxic sample 1, 14,226,458 from hypoxic sample 2, 14,697,154 from hypoxic sample 3, 14,406,854 from normoxic sample 1, 14,988,132 from normoxic sample 2, and 11,516,478 from normoxic sample 3 were referenced to the paired-end match list to generate genome-wide enhancer activity tables by 100bp bins. The genomic fragments ranged from 400-500bp so the bin corresponding to the alignment as well as the four downstream bins were credited 1 count. In the cases where randomer tags matched multiple genomic fragments, bins were credited a fraction of a count based on the likelihood of that linkage in the paired-end match data. This created a genome-wide count table of enhancer activity in each replicate.

The count table was then analyzed in R for differential activity between hypoxic and normoxic replicates using a negative binomial test in the DESeq[29] package. The bins were filtered by overall count ($\theta=0.5$) and the test was run with default variance estimation. This generated a p-value and a p-value adjusted for multiple hypothesis testing (Benjamini-Hochberg procedure) for each 100bp bin. Hypoxic enhancer regions were defined at bins up-regulated under hypoxia with adjusted $p\text{-value} < 0.1$ ($p\text{-value} < 1.55\text{ e-}05$) and extend to include adjacent bins with $p\text{-value} < 0.05$.

In order to compare our results to those of STARR-seq[20], statistically significant S2 baseline enhancers were identified in the three normoxic replicates. In this case, R was used to perform a negative binomial test between counts in the DNA library and counts in the RNA-sequenced barcodes for each 100bp bin across the *Drosophila melanogaster* genome in order to identify genomic regions enriched for enhancer activity. Peaks were identified with an adjusted p-value < 0.018 (Benjamini-Hochberg procedure) and enhancers were defined as the 500 bp interval surrounding the activity peak in order to maintain consistency with STARR-seq data.

Enhancer sequence motif analysis

Identified hypoxic enhancer regions were searched for stress transcription factor binding sites using the BoBro BBS motif-scanning algorithm[30] with position weight matrices from the JASPAR database[31]. This algorithm was used to identify binding site positions and calculate a global p-value of enrichment for HIF-1 (JASPAR ID: MA0259.1), FOXO (MA0480.1), HSF (MA0486.1) and NF-kB (MA0105.3) binding sites in enhancer sequences compared to the *Drosophila melanogaster* genome background.

RESULTS

Discovered hypoxic enhancers

Transcriptional activity from 4,599,881 fragments that were 400-500bp in size, spanning the *Drosophila melanogaster* genome at 17.39X coverage, was analyzed by 100bp bins and 31 significant hypoxic enhancer regions (q-value < 0.1, p-value < 1.55 e-05) were identified (Table 1, Supplementary File S5). These enhancer regions range in size from 100 to 800bp and confer 2 to 18-fold changes in expression under hypoxia. The discovered enhancers are found throughout the genome and are located proximally to genes up-regulated under hypoxia in our RNAseq experiments. The ten most strongly up-regulated genes all contain a discovered enhancer within 20kb. 16 of 31 discovered enhancers are located within 20kb of one of the 90 up-regulated genes. The probability of this positional overlap occurring by chance is 1.43 e-14 using an exact binomial test, supporting that the discovered enhancers are linked to endogenous gene expression and implicating their likely targets. 4 additional enhancers are proximal to genes previously observed to be up-regulated under hypoxia in *Drosophila*[32].

Location of hypoxic enhancers

Of the 20 hypoxic enhancer regions proximal (within 20kb) to hypoxic up-regulated genes, 6 fall in the promoter region of the putative target gene (Figure 2, Table 1). All six of these are the homologous Hsp70B enhancers. Six enhancers were found in introns of putative target genes (Table 1). These intronic enhancers may be placed proximal to alternate transcription start sites in order to confer isoform specific up-regulation as seen in the case of Sima, the *Drosophila* HIF-1α homologue (Figure 3). Two enhancers were found in introns of genes neighbouring the putative target and one was found in the ORF of the putative target. The remaining five were found in intergenic space up or downstream of putative target genes, as seen for the enhancer region 13 kb downstream of the transcriptional regulator hairy (Figure 4). Interestingly, three of the five intergenic enhancers were located immediately proximal to a ncRNA. All of these ncRNAs were themselves up-regulated under hypoxia (Table 2).

Transcription factor binding motifs

Identified enhancer regions are enriched for binding sites of stress response transcription factors involved in hypoxia. Transcription factors HSF, HIF-1, FOXO, and NF-kB showed highly significant global enrichment across the enhancer regions (Table 3). Binding sites occurring in each individual enhancer are listed in Table 1. 26 of 31 enhancer regions contain binding motifs for at least one of these transcription factors and many contain binding sites for several. In addition to a pair of HSF binding sites, The Hsp70B promoter proximal enhancers contain binding sites for FOXO and HIF-1 (Figure 2). The intronic Sima enhancer (Figure 3) contains a pair of HIF-1 binding sites, possibly allowing autoregulation, and also contains a NF-kB binding site. The enhancer region downstream of hairy contains HSF, FOXO, and HIF-1 binding sites (Figure 4).

Overlap with STARR-seq enhancers

Our data correlate strongly with a previous genome-wide empirical assay of Drosophila transcriptional enhancers. STARR-seq[20] was used to identify 5499 enhancers operating in S2 cells under normal conditions. These enhancers were defined as 500 bp intervals surrounding statistically significant peaks in enhancer activity (Adjusted p-value < 0.018, p < 0.001). In order to generate a similar dataset for comparison, we identified genomic regions showing significant enrichment in normoxic S2 cells. Similar to STARR-seq, we defined enhancers as 500 bp intervals surrounding peaks with an adjusted p-value less than 0.018 (unadjusted < 0.00043). This yielded a list of 1007 baseline S2 enhancers (Supplementary File S6). 466 of these (46.3%) overlap the enhancers identified using STARR-seq. The probability of a 500 bp fragment overlapping the STARR-seq set by chance is 0.0462. An exact binomial test (463 hits, 1007 trials, 0.0462 background probability) generates a p-value of 0 for the overlap between our data and STARR-seq. The ratio of overlap is higher when only the most enriched peaks from out dataset are examined. 19 of the 25 most enriched enhancer peaks (76%, p-value = 5.66e-21) overlap with the STARR-seq dataset. This high degree of overlap demonstrates a robust ability to identify active enhancers.

DISCUSSION

We used a novel parallelized reporter assay to conduct the first genome-wide functional enhancer screen of a cellular response to environmental stress. Our work demonstrates a new method with wide applicability and identifies DNA regulatory sequences conferring hypoxic activity. We identify 31 hypoxic enhancer regions and analyze them with respect to up-regulated hypoxic genes and stress response transcription factors.

RNA-Seq was performed on the same RNA pools used to quantify hypoxic enhancer activity in order to identify putative target genes proximal to identified enhancer regions. Differentially expressed genes identified in our RNA-Seq experiments are corroborated by previous analyses of the *Drosophila* hypoxic response[32,33]. The majority of enhancer regions were proximal (within 20 kb) to endogenously up-regulated genes, indicating that our enhancer assay identifies active *in vivo* regulatory elements. We identified enhancer regions proximal to previously described hypoxic genes including lactate dehydrogenase[6,32], the transcriptional regulator hairy[34], the reductase Wwox[35], and the cell cycle inhibitor scyl[36]. Additionally, the Hsp70B promoter proximal enhancers identified in our assay have been previously shown to be active *in vivo*[37,38]. The large positional overlap between up-regulated genes and enhancer regions allowed analysis of the architecture of hypoxic regulation. Interestingly, while some enhancers were found at the promoters of putative target genes, the majority of enhancer regions were found in introns and intergenic space. Enhancers were found in introns of putative target genes as well as introns of neighboring genes (Table 1). Enhancer regions in intergenic space corresponded with known ncRNA loci and in each case the ncRNA was itself up-regulated under hypoxia (Table 2). These findings highlight the unbiased view of the regulatory landscape provided by genome-wide empirical assays and underscore the prevalence of activity outside of promoter regions.

Some of the enhancer regions were not proximal to an identifiable up-regulated gene. These enhancers could act on more distal targets, on proximal targets with expression too low to be detected by our RNA-Seq experiment, or they may have activity in isolation but be attenuated by other elements in their native hypoxic context. Conversely, many up-regulated genes did not have a proximal enhancer identified by our screen. This could be due to a requirement of action from multiple disjunct regulatory modules at the native locus or lack of resolution in our assay. The resolution of our assay was attenuated by the coincidence of randomer tags with multiple genomic regions. Other randomer tag-based approaches test orders of magnitude fewer fragments and hence largely avoid barcode collision[16-19]. This problem is circumvented in STARR-Seq[20] by confining placement of the potential enhancer to the transcribed region so that it can be assayed directly by RNA sequencing. Future uses of our technique will benefit from further optimization of library synthesis to allow a greater number of randomer tags into the library. Nonetheless, the technique is highly functional in its present state and introduces a simpler and more versatile library synthesis approach. Indeed, our data show a large degree of overlap with STARR-Seq with respect to baseline transcriptional enhancers in *Drosophila* S2 cells. Furthermore, this work presents a large list of empirically identified hypoxia-induced enhancer regions robust to false discovery rate that coincide with the most highly up-regulated hypoxic genes.

The transcription factors HIF-1, HSF, NF-kB , and FOXO regulate hypoxic gene expression and have been shown to exhibit overlapping activity and reciprocal regulation[9-11,39,40]. The enhancer regions identified in this study are highly enriched for their binding site motifs and many display multiple sites allowing signal integration of stress response pathways. We observe an intronic enhancer in Sima which contains both HIF-1 and NF-kB binding sites, suggesting HIF-1 autoregulation and integration of NF-kB signaling at a basal level in the hypoxic response. The enhancer region, while intronic to the full-length Sima transcript isoforms, is upstream of an alternative transcriptional start site that produces a transcript isoform that is up-regulated after hypoxia, whereas the full-length isoforms do not have altered expression after hypoxic stress. This short isoform lacks the bHLH and PAS domains of the full-length isoform, suggesting it neither binds DNA nor heterodimerizes. Interestingly, this hypoxic regulation of a short isoform resembles the hypoxic induction of a short isoform of the HIF-1 regulator *fatiga* (Drosophila HIF-1 Prolyl Hydroxylase) by an intronic HIF-1 enhancer[41].

Our findings reiterate the complexities of the hypoxic response while providing new details. The enhancer regions identified demonstrate regulatory activity distributed throughout non-coding genomic space and underscore the role of intronic enhancers in the hypoxic response. We observe coincidence between enhancer regions and ncRNA activity in agreement with previous evidence showing local transcription to be a general property of active enhancers[42]. We present a set of sequences capable of driving hypoxia-specific expression and demonstrate a new genome-wide technique for the identification of context-specific enhancers.

ACKNOWLEDGEMENT

We thank Paul Etter and Doug Turnbull for advice on Illumina library preparation.

REFERENCES

1. Bulger,M. and Groudine,M. (2011) Functional and mechanistic diversity of distal transcription enhancers. *Cell,* **44**, 327-339.
2. Perry,M.W., Boettiger,A.N. and Levine,M. (2011) Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. *Proc. Natl. Acad. Sci. USA,* **108**, 13570-13575.
3. Lagha,M., Bothma,J.P. and Levine,M. (2012) Mechanisms of transcriptional precision in animal development. *Trends Genet*., **28**, 409-416.
4. Arnosti,D.N. and Kulkarni,M.M. (2005) Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards?. *J. Cell. Biochem*., **94**, 890-898.
5. Swanson,C., Schwimmer,D.B. and Barolo,S. (2011) Rapid evolutionary rewiring of a structurally constrained eye enhancer. *Curr. Biol*., **21**, 1186-1196.
6. Bruick,R.K. and McKnight,S.L. (2001) A conserved family of prolyl-4-hydroxylases that modify HIF. *Sci. Signal*., **294**, 1337-1340.
7. Lavista-Llanos,S., Centanin,L., Irisarri,M., Russo,D.M., Gleadle,J.M., Bocca,S.N., Muzzopappa,M., Ratcliffe,P.J. and Wappner,P. (2002) Control of the hypoxic response in Drosophila melanogaster by the basic helix-loop-helix PAS protein similar. *Mol. Cell Biol*., **22**, 6842-6853.
8. Wang,G.L. and Semenza,G.L. (1993) General involvement of hypoxia-inducible factor 1 in transcriptional response to hypoxia. *Proc. Natl. Acad. Sci. USA*, **90**, 4304–4308.
9. Rius,J., Guma,M., Schachtrup,C., Akassoglou,K., Zinkernagel,A.S., Nizet,V., Johnson,R.S., Haddad,G.G. and Karin,M. (2008) NF-kappaB links innate immunity to the hypoxic response through transcriptional regulation of HIF-1alpha. *Nature*, **453**, 807–811.
10. van Uden,P., Kenneth,N.S., Webster,R., Müller,H.A., Mudie,S, and Rocha,S. (2011) Evolutionary Conserved Regulation of HIF-1β by NF-κB. *PLoS Genet*., **7**, e1001285.
11. Scortegagna,M., Cataisson,C., Martin,R.J., Hicklin,D.J., Schreiber,R.D., Yuspa,S.H. and Arbeit,J.M. (2008) HIF-1α regulates epithelial inflammation by cell autonomous NFκB activation and paracrine stromal remodeling. *Blood*, **111**, 3343-3354.
12. Metzker,M.L. (2010) Sequencing technologies—the next generation. *Nat. Rev. Genet*., **11**, 31-46.

13. Wang,Z., Gerstein,M. and Snyder,M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.*, **10**, 57-63.

14. Johnson,D., Mortazavi,A., Myers,R.M. and Wold,B. (2007) Genome-wide mapping of in vivo protein-DNA interactions. *Science*, **316**, 1497-1502.

15. Song,L., Zhang,Z., Grasfeder,L.L., Boyle,A.P., Giresi,P.G., Lee,B.K., Sheffield,N.C., Graf,S., Huss,M., Keefe,D. et al. (2011) Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome research,* **21**,1757-1767.

16. Kwasnieski,J.C., Mogno,I., Myers,C.A., Corbo,J.C. and Cohen,B.A. (2012) Complex effects of nucleotide variants in a mammalian cis-regulatory element. *Proc. Natl. Acad. Sci. USA*, **109,** 19498-19503.

17. Patwardhan,R.P., Hiatt,J.B., Witten,D.M., Kim,M.J., Smith,R.P., May,D. Lee,C., Andrie,J.M., Lee,S., Cooper,G.M., Ahituv,N., Pennacchio,L.A. and Shendure,J. (2012) Massively parallel functional dissection of mammalian enhancers in vivo. *Nat. Biotechnol.*, **30**, 265-270.

18. Melnikov A, Murugan A, Zhang X, Tesileanu T, Wang L, Rogov,P., Feizi,S., Gnirke,.A, Callan,C.G., Kinney,J.B., Kellis,M., Lander,E.S. and Mikkelsen,T.S. (2012) Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.*, **30**, 271-277.

19. Kheradpour P, Ernst J, Melnikov A, Rogov P, Wang L, Zhang,X., Alston,J., Mikkelsen,T.S. and Kellis,M. (2013) Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res.*, **23**, 800-811.

20. Arnold,C.D., Gerlach,D., Stelzer,C., Boryń,Ł.M., Rath,M. and Stark,A. (2013) Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. *Science*, **339**, 1074-1077.

21. Bulger,M. and Groudine,M. (2011) Functional and mechanistic diversity of distal transcription enhancers. *Cell,* **44**, 327-339.

22. Banerji,J., Rusconi.,S and Schaffner,W. (1981) Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell,* **27**, 299-308.

23. Amit,R., Garcia,H.G., Phillips,R. and Fraser,S.E. (2011) Building enhancers from the ground up: a synthetic biology approach. *Cell*, **146**, 105-118.

24. Erokhin,M., Davydova,A., Parshikov,A., Studitsky,V.M., Georgiev,P. and Chetverina,D. (2013) Transcription through enhancers suppresses their activity in Drosophila. *Epigenetics Chromatin,* **6**, 31.

25. Zieler,H. and Huynh,C.Q. (2002) Intron-dependent stimulation of marker gene expression in cultured insect cells. *Insect Mol. Biol*., **11**, 87-95.

26. Pfeiffer,B.D., Jenett,A., Hammonds,A.S., Ngo,T.T.B., Misra,S., Murphy,C., Scully,A., Carlson,J.W., Wan,K.H., Laverty,T.R., Mungall,C., Svirskas,R., Kadonaga,J.T., Doe,C.Q., Eisen,M.B., Celniker,S.E. and Rubin,G.M. (2008) Tools for neuroanatomy and neurogenetics in Drosophila. *Proc. Natl. Acad. Sci. USA*, **105**, 9715-9720.

27. Trapnell,C., Lior,P. and Salzberg,S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics,* **25**, 1105-1111.

28. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods,* **9**, 357-359.

29. Anders,S. and Huber,W. (2010) Differential expression analysis for sequence count data. *Genome Biol.,* **11**, R106.

30. Ma,Q., Zhang,H., Mao,X., Zhou,C., Liu,B., Chen,X., and Xu,Y. (2014) DMINDA: an integrated web server for DNA motif identification and analyses. *Nucleic Acids Res.*, gku315.

31. Mathelier,A., Zhao,X., Zhang,A.W., Parcy,F., Worsley-Hunt,R., Arenillas,D.J., Buchman,S., Chen,C., Chou,A., Ienasescu,H., Lim,J., Shyr,C., Tan,G., Zhou,M., Lenhard,B., Sandelin,A. and Wasserman,W.W. (2013) "JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles." *Nucleic Acids Res*, gkt997.

32. Li,Y., Padmanabha,D., Gentile,L.B., Dumur,C.I., Beckstead,R.B. and Baker,K.D. (2013) HIF-and Non-HIF-Regulated Hypoxic Responses Require the Estrogen-Related Receptor in Drosophila melanogaster. *PLoS Genet*, **9**, e1003230.

33. Liu,G., Roy,J. and Johnson,E.A. (2006) Identification and function of hypoxia-response genes in Drosophila melanogaster. *Physiol. Genomics,* **25**, 134-141.

34. Zhou,D., Xue,J., Lai,J.C., Schork,N.J., White,K.P. and Haddad,G.G. (2008) Mechanisms underlying hypoxia tolerance in Drosophila melanogaster: hairy as a metabolic switch. *PLoS Genet*., **4**, e1000221.

35. O'Keefe,L.V., Colella,A., Dayan,S., Chen,Q., Choo,A., Jacob,R., Price,G., Venter,D., and Richards,R.I. (2011) Drosophila orthologue of WWOX, the chromosomal fragile site FRA16D tumour suppressor gene, functions in aerobic metabolism and regulates reactive oxygen species. *Hum. Mol. Genet*., **20**, 497-509.

36. Scuderi,A., Simin,K., Kazuko,S.G., Metherall,J.E. and Letsou,A. (2006) scylla and charybde, homologues of the human apoptotic gene RTP801, are required for head involution in Drosophila. *Dev Biol*, **291**, 110-122.

37. Tian,S., Haney,R. and Feder,M. (2010) Phylogeny disambiguates the evolution of heat-shock cis-regulatory elements in Drosophila. *PLoS One*, **5**, e10669.

38. Li,D., Li,G., Wang,K., Liu,X., Li,W., Chen,X. and Wang,Y. (2012) Isolation and functional analysis of the promoter of the amphioxus Hsp70a. *Gene*, **510**, 39-46.

39. Hsu,A.L., Murphy,C.T. and Kenyon,C. (2003) Regulation of aging and age-related disease by DAF-16 and heat-shock factor. *Science*, **300**, 1142-1145.

40. Wang,M.C., Bohmann,D. and Jasper,H. (2005) JNK extends life span and limits growth by antagonizing cellular and organism-wide responses to insulin signaling. *Cell*, **12**, 115-125.

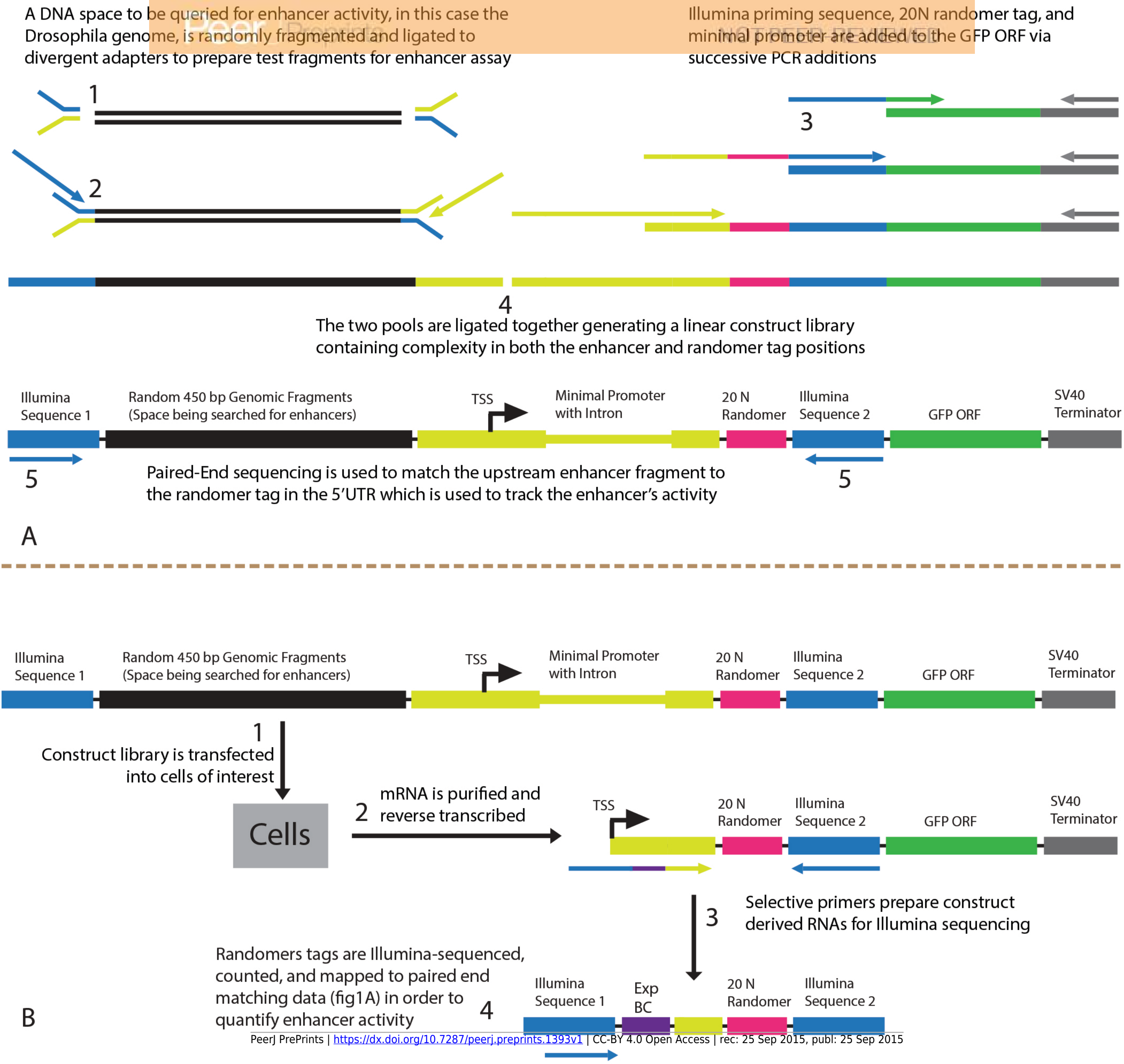
41. Acevedo,J.M., Centanin,L., Dekanty,A. and Wappner,P. (2010) Oxygen Sensing in *Drosophila*: Multiple Isoforms of the Prolyl Hydroxylase Fatiga Have Different Capacity to Regulate HIFα/Sima. *PLoS One*, **5**, e12390.

42. Andersson,R., Gebhard,C., Miguel-Escalada,I., Hoof,I., Bornholdt,J., Boyd,M., Chen,Y., Zhao,X., Schmidl,C., Suzuki,T., et al. (2014) An atlas of active enhancers across human cell types and tissues. *Nature*, **507**, 455-461.

Figure 1(on next page)

Library Synthesis

(A) The enhancer library is synthesized entirely *in vitro*. DNA of interest is fragmented (step 1) and ligated to divergent adapters (step 2) leaving potential enhancer fragments with Illumina sequence on one side and the beginning of the synthetic minimal promoter on the other. The GFP gene is used as a template for a series of 5’ PCR additions in order to add Illumina sequence, 20N randomer tag, and the majority of the minimal promoter and intron (step 3). The two sides are ligated together to create a linear construct with complexity in the enhancer region upstream of the transcription start site as well as complexity in the randomer tag region in the 5’ UTR (step 4). The sample is submitted to paired-end sequencing in order to match the potential enhancer region to the randomer tag in the 5’ UTR that is used to report its activity. **(B)** The enhancer library is transfected into cells (step 1) and total RNA is purified and reverse transcribed to create cDNA (step 2). The cDNA is used as template for a PCR reaction (step 3) with a reverse primer complimentary to the Illumina end 2 sequence present in the construct and a forward primer complimentary to the stretch of the minimal promoter upstream of the randomer tag. The forward primer adds Illumina end 1 seqeunce and an experimental barcode for multiplexing. This amplicon is ready to be loaded onto the Illumina flow cell for single-end sequencing of randomer tags (step 4) in order to quantify enhancer activity.



A DNA space to be queried for enhancer activity, in this case the Drosophila genome, is randomly fragmented and ligated to divergent adapters to prepare test fragments for enhancer assay

Illumina priming sequence, 20N randomer tag, and minimal promoter are added to the GFP ORF via successive PCR additions

The two pools are ligated together generating a linear construct library containing complexity in both the enhancer and randomer tag positions

Paired-End sequencing is used to match the upstream enhancer fragment to the randomer tag in the 5'UTR which is used to track the enhancer's activity

Construct library is transfected into cells of interest

mRNA is purified and reverse transcribed

Selective primers prepare construct derived RNAs for Illumina sequencing

Randomers tags are Illumina-sequenced, counted, and mapped to paired end matching data (fig1A) in order to quantify enhancer activity

Figure 2(on next page)

Hypoxic Enhancer Activity by 100bp Bins at the hairy Locus

Each open circle plots the p-value of the difference in randomer tag counts mapping to that 100bp bin between normoxia and hypoxia. The green bar shows the enhancer region discovered by our genome-wide screen. **(A)** The hairy gene produces a negative transcriptional regulator that is up-regulated during hypoxia. We identify an active hypoxic enhancer 13kb downstream of hairy. **(B)** The close up of the hairy downstream enhancer region shows FOXO, HIF-1 and HSF binding sites as well as coincidence with a ncRNA that is also up-regulated under hypoxia.

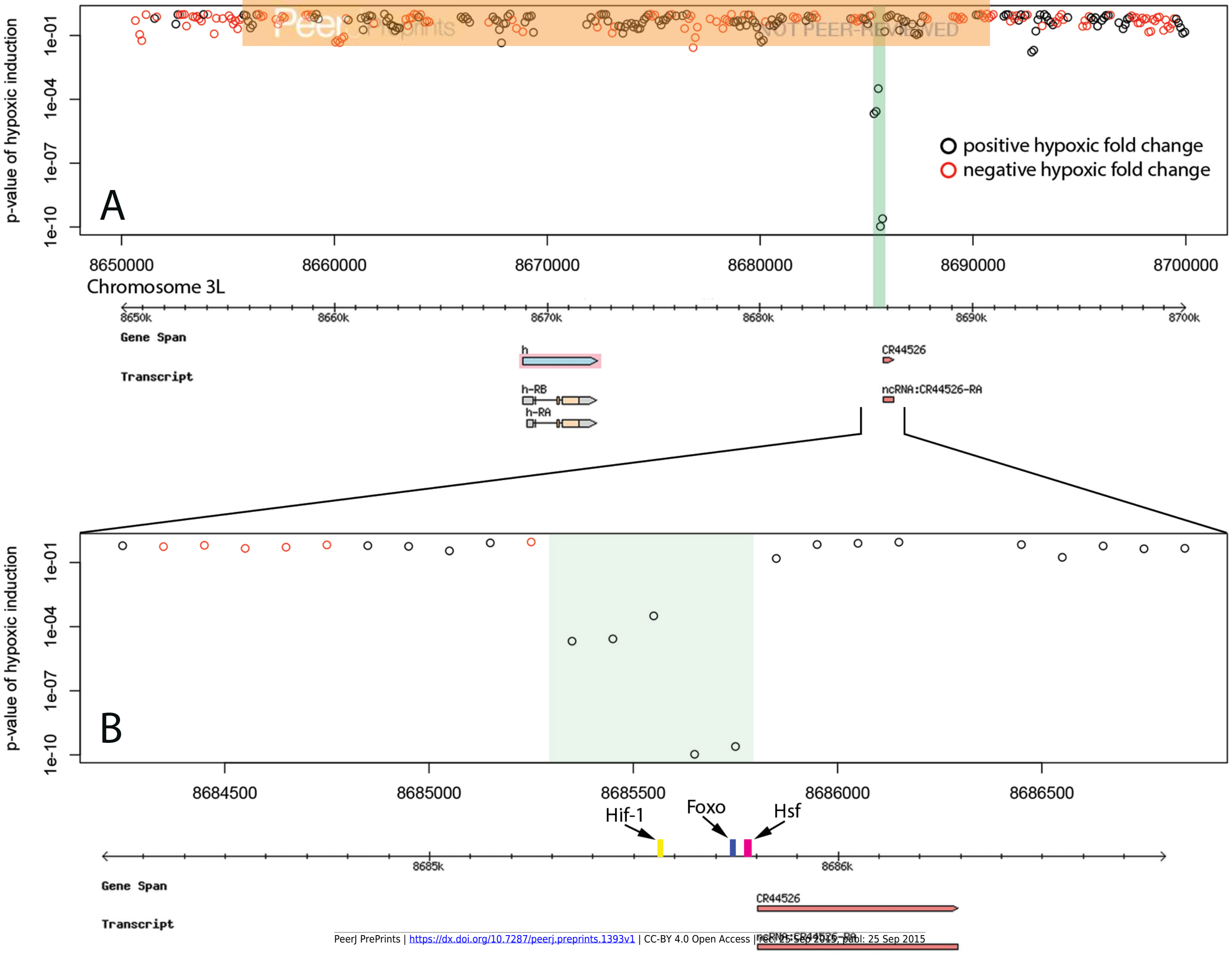

A
p-value of hypoxic induction
1e-01
1e-04
1e-07
1e-10
positive hypoxic fold change
negative hypoxic fold change
8650000
8660000
8670000
8680000
8690000
8700000
Chromosome 3L
8650k
8660k
8670k
8680k
8690k
8700k
Gene Span
Transcript
h
h-RB
h-RA
CR44526
ncRNA:CR44526-RA
B
p-value of hypoxic induction
1e-01
1e-04
1e-07
1e-10
8684500
8685000
8685500
8686000
8686500
Hif-1
Foxo
Hsf
8685k
8686k
Gene Span
Transcript
CR44526
ncRNA:CR44526-RA

Figure 3(on next page)

Hypoxic Enhancer Activity by 100bp Bins at the Hsp70B Locus

Each open circle plots the p-value of the difference in randomer tag counts mapping to that 100bp bin between normoxia and hypoxia. Green bars show enhancer regions discovered by our genome-wide screen. **(A)** The four Hsp70B homologues highlighted in pink are all up-regulated under hypoxia and contain homologous promoter proximal hypoxic enhancer regions. Additionally, a fifth homologous enhancer region lacking an ORF was discovered at the locus. **(B)** The close up of the Hsp70Ba enhancer region shows the position of multiple stress response transcription factor binding sites.

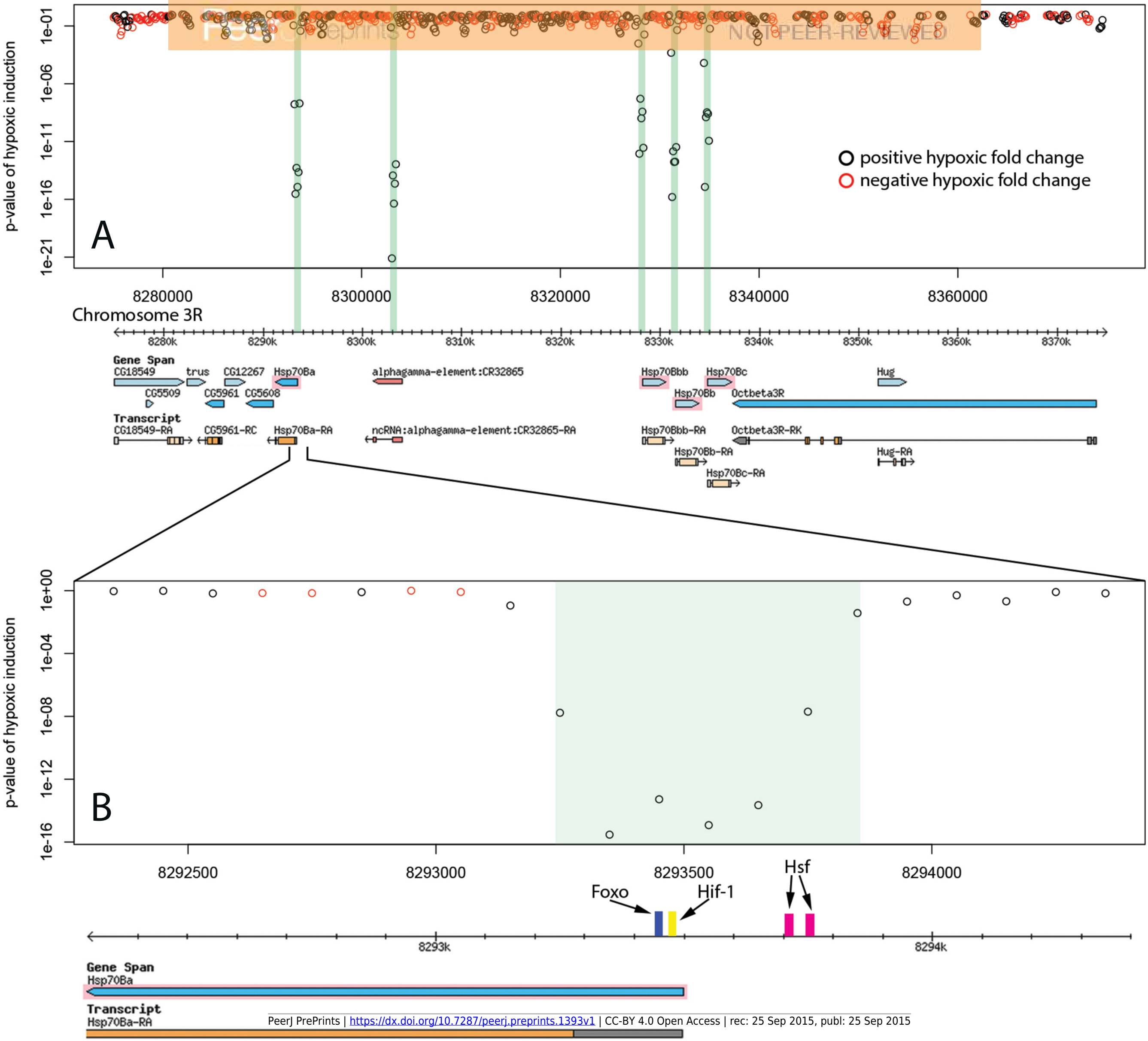

A
p-value of hypoxic induction
1e-01
1e-06
1e-11
1e-16
1e-21
positive hypoxic fold change
negative hypoxic fold change
8280000
8300000
8320000
8340000
8360000
Chromosome 3R
8280k
8290k
8300k
8310k
8320k
8330k
8340k
8350k
8360k
8370k
Gene Span
CG18549
trus
CG12267
Hsp70Ba
alphagamma-element:CR32865
Hsp70Bbb
Hsp70Bc
Hug
CG5509
CG5961
CG5608
Hsp70Bb
Octbeta3R
Transcript
CG18549-RA
CG5961-RC
Hsp70Ba-RA
ncRNA:alphagamma-element:CR32865-RA
Hsp70Bbb-RA
Octbeta3R-RK
Hsp70Bb-RA
Hug-RA
Hsp70Bc-RA
B
p-value of hypoxic induction
1e+00
1e-04
1e-08
1e-12
1e-16
8292500
8293000
8293500
8294000
Foxo
Hif-1
Hsf
8293k
8294k
Gene Span
Hsp70Ba
Transcript
Hsp70Ba-RA

Figure 4(on next page)

Hypoxic Enhancer Activity by 100bp Bins at the Sima (HIF-1α) Locus

Each open circle plots the p-value of the difference in randomer tag counts mapping to that 100bp bin between normoxia and hypoxia. The green bar shows the enhancer region discovered by our genome-wide screen. **(A)** HIF-1 is the master hypoxic regulator and is itself regulated transcriptionally under hypoxia. Our RNASeq data shows hypoxia induces up-regulation of the isoform highlighted in pink. We identify an intronic hypoxic enhancer upstream of the transcription start site of this isoform. **(B)** The close up of the Sima intronic enhancer region shows both HIF-1 and NF-kB binding sites.

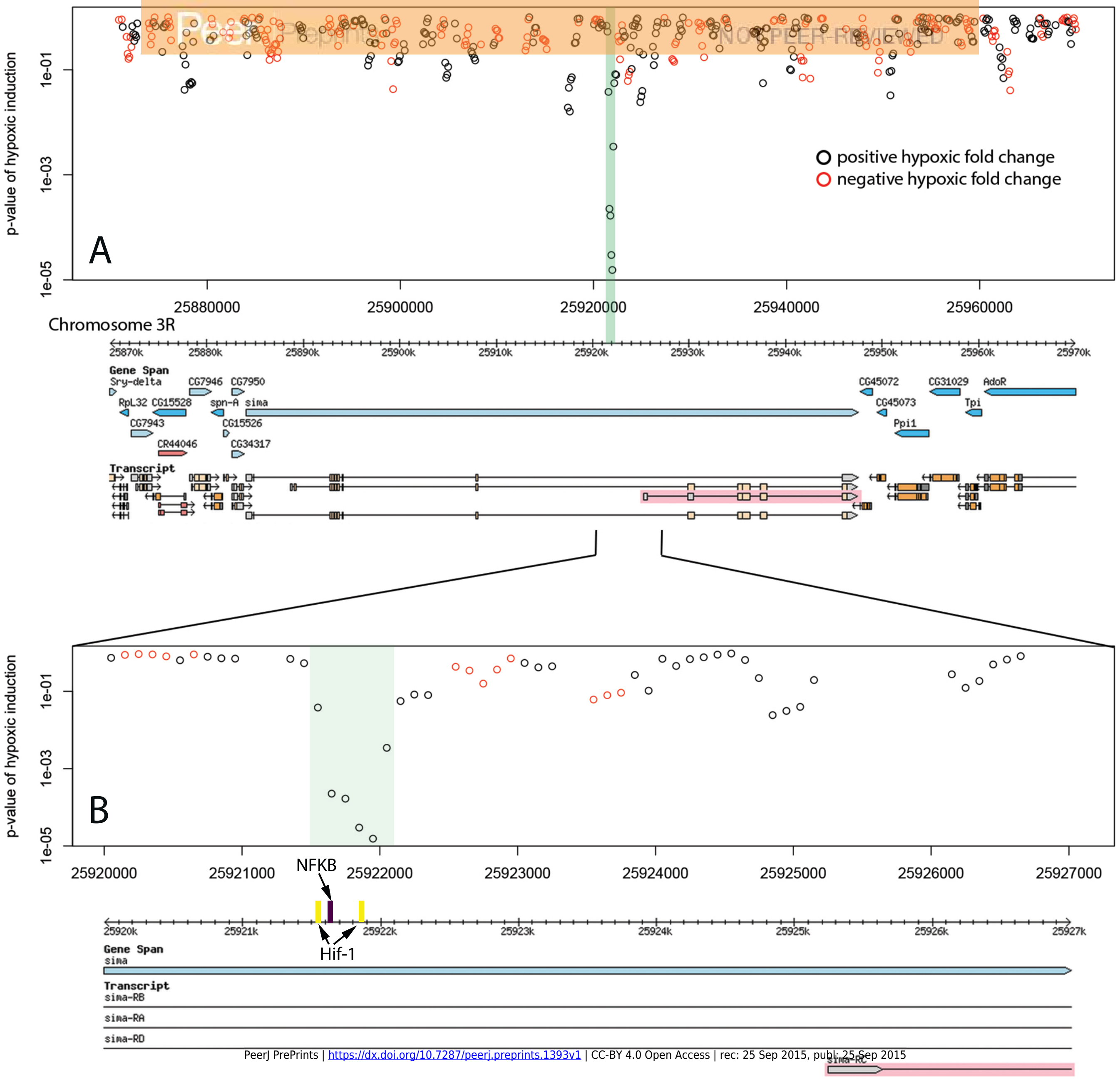

A
p-value of hypoxic induction
1e-01
1e-03
1e-05
positive hypoxic fold change
negative hypoxic fold change
25880000
25900000
25920000
25940000
25960000
Chromosome 3R
25870k
25880k
25890k
25900k
25910k
25920k
25930k
25940k
25950k
25960k
25970k
Gene Span
Sry-delta
CG7946
CG7950
RpL32
CG15528
spn-A
sima
CG7943
CG15526
CR44046
CG34317
CG45072
CG31029
AdoR
CG45073
Tpi
Ppi1
Transcript
B
p-value of hypoxic induction
1e-01
1e-03
1e-05
25920000
25921000
25922000
25923000
25924000
25925000
25926000
25927000
NFKB
Hif-1
25920k
25921k
25922k
25923k
25924k
25925k
25926k
25927k
Gene Span
sima
Transcript
sima-RB
sima-RA
sima-RD
sima-RC

Table 1(on next page)

Properties of Discovered Hypoxic Enhancers

The 31 hypoxic enhancers identified by our genome-wide screen are shown in order of statistical significance. Column one is the genomic location of the enhancer (dm3). Column two is the p-value between hypoxic and normoxic counts as calculated by the negative binomial test with column three showing the Benjamini-Hochberg adjusted p-value. Column four is the fold change of transcriptional activity due to the enhancer in hypoxic versus normoxic conditions. Column five shows endogenous genes within 20kb that were significantly up-regulated under hypoxia in the same RNA extracts used to calculate enhancer activity. The rank of the gene's hypoxic induction is shown in parentheses and genes marked with an asterisk were observed to be up-regulated under hypoxia in *Drosophila* by Li et al.[27]. Column six indicates the relative position of the enhancer to the proximal hypoxic up-regulated gene. Column seven shows binding sites for stress-related transcription factors found in the enhancer.

Enhancer Locus	P-value	Adjusted P-value	Fold Change	Hypoxic Gene(s) Within 20Kb	Relative Position to Hypoxic Gene(s)	Stress TF Binding Sites
3R:8303000..8303500	7.79 e-22	4.63 e-16	5.08	Hsp70B genes (1-4)	Intergenic	Hsf, Hif-1, Foxo
3L:6256700..6257200	1.83 e-16	2.72 e-11	5.95	impl3 (9)	Upstream	NF-kB
3R:8331100..8331800	1.59 e-16	2.72 e-11	4.49	Hsp70Bb (2)	Promoter Proximal	Hsf, Hif-1, Foxo
3R:8293200..8293900	2.96 e-16	3.51 e-11	3.83	Hsp70Ba (4)	Promoter Proximal	Hsf, Hif-1, Foxo
3R:8334400..8335000	1.18 e-15	1.01 e-10	4.45	Hsp70Bc (1)	Promoter Proximal	Hsf, Hif-1, Foxo
2L:8001300..8001800	2.64 e-15	1.74 e-10	6.44	Wwox (15)	Intronic	Hif-1
3R:8327800..8328500	8.89 e-13	2.40 e-08	3.70	Hsp70Bbb (3)	Promoter Proximal	Hsf, Hif-1, Foxo
2L:20082900..20083500	1.08 e-12	2.79 e-08	6.35	Fok (11)	Intronic	Foxo, Hif-1
3L:8685300..8685800	1.07 e-10	2.18 e-06	3.79	Hairy (45)	Downstream	Hsf, Hif-1, Foxo
3L:7797800..7798600	1.77 e-10	3.38 e-06	3.07	CG32369 (23)	Intronic	Hif-1
3L:9385200..9385800	2.14 e-09	3.62 e-05	3.71	Hsp22,23,26,27 (7,8,10,14)	Neighboring Intron	Not Detected
X:17071000..17071300	8.77 e-09	1.24 e-04	4.99	Not Detected	Not Detected	Not Detected
X:9767000..9767500	1.27 e-08	1.76 e-04	3.65	CG32695*	ORF	Not Detected
2L:2887100..2887600	1.32 e-08	1.79 e-04	5.82	Not Detected	Not Detected	Hif-1
3L:11234100..11234900	6.03 e-07	6.63 e-03	2.68	Scylla (19)	Upstream	Foxo
3L:3892900..3893100	1.55 e-06	1.59 e-02	2.75	Not Detected	Not Detected	Hif-1, NF-kB
2L:5986900..5987500	1.82 e-06	1.81 e-02	2.16	ifc*	Intronic	Foxo
3L:9448800..9448900	2.09 e-06	2.03 e-02	5.39	MTF-1*	Neighboring Intron	NF-kB, Hif-1
3R:6800900..6801600	2.22 e-06	2.09 e-02	13.82	Not Detected	Not Detected	Hif-1
3L:11522800..11523300	2.66 e-06	2.35 e-02	3.04	Not Detected	Not Detected	NF-kB
3R:4181100..4181600	2.66 e-06	2.35 e-02	3.87	Atg13 (51)	Downstream	Foxo, Hif-1
3R:7781900..7782700	2.69 e-06	2.35 e-02	4.96	Hsp70Aa (6)	Promoter Proximal	Hsf
3R:7783900..7784500	2.75 e-06	2.37 e-02	4.18	Hsp70Ab (5)	Promoter Proximal	Hsf
3R:21433600..21434000	3.30 e-06	2.72 e-02	9.03	Not Detected	Not Detected.	Not Detected
X:16559200..16559700	4.13 e-06	3.23 e-02	6.56	Not Detected	Not Detected	Foxo
3R:2902300..2902600	6.21 e-06	4.63 e-02	2.95	Not Detected	Not Detected	Not Detected
2R:12896000..12896500	6.88 e-06	5.05 e-02	3.02	Not Detected	Not Detected	Foxo
X:17388000..17388500	8.24 e-06	5.75 e-02	6.80	Not Detected	Not Detected.	Hif-1
3R:14892300..14892800	9.76 e-06	6.44 e-02	18.01	Not Detected	Not Detected	Hif-1
3R:27050000..27050500	1.52 e-05	9.40 e-02	2.78	CG12054*	Intronic	Hif-1
3R:25921500..25922100	1.54 e-05	9.44 e-02	2.46	Hif-1 (71)	Intronic	NF-kB, Hif-1

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1393v1 | CC-BY 4.0 Open Access | rec: 25 Sep 2015, publ: 25 Sep 2015

Table 2(on next page)

P-value of Stress Transcription Factor Binding Site Enrichment in Discovered Enhancer Sequences

Transcription Factor	**P-value of Enrichment**
HSF	6.22 e-12
Hif-1	6.49 e-06
Foxo	1.01 e-04
NF-kB	6.67 e-04

Table 3(on next page)

ncRNAs Proximal to Hypoxic Enhancers

Three of the five enhancers not contained within protein coding transcripts coincide with ncRNAs. Each of these ncRNAs is also up-regulated under hypoxia.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1393v1 | CC-BY 4.0 Open Access | rec: 25 Sep 2015, publ: 25 Sep 2015

Enhancer Locus	**ncRNA**	**Position of ncRNA relative to enhancer**	**Hypoxic read counts**	**Normoxic read counts**
3R:8303000..8303500	CR32865	overlapping	66	13
3L:8685300..8685800	CR44526	3 bp upstream	31	14
3L:6256700..6257200	CR44522	201 bp upstream	6	1

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1393v1 | CC-BY 4.0 Open Access | rec: 25 Sep 2015, publ: 25 Sep 2015