Interpretable Machine Learning Approach - Elucidate Messenger RNA Subcellular Localization Codes

COMP 680: Mining Biological Sequences — Mathieu Blanchette, Winter 2017

Mansha Imtiyaz (260712985)
mansha.imtiyaz@mail.mcgill.ca

Atefeh Mohajeri
at.mohajeri@gmail.com

***Abstract*—In this report we introduce the work of Ribeiro *et al.* [1] which presents a model-agnostic interpretation tool called LIME. LIME enables us to explain individual predictions of a Machine Learning technique by learning an *interpretable model* locally around the prediction. We develop a pipeline for predicting the localization patterns of mRNAs in Drosophila melanogaster embryos, and use LIME to identify RBPs involved in this process.**

I. Introduction

Subcellular localization of mRNAs have been known as a mechanism for locally producing proteins. Nevertheless, its only in recent studies that subcellular localization has emerged as the main mechanism underlying the localization of proteins in many polarized cells. Lecuyer et al [2] has discovered that in Drosophila embryo, more than 70% of the 2314 expressed transcripts were sub cellularly localized. All fundamental developmental cell processes such as asymmetric cell division or embryonic patterning highly depend on cell polarization, which in turn depends on cytoplasmic mRNA localization.

There are several reasons why transporting mRNA might be preferable to transporting proteins. It reduces the cost of transport since one mRNA can produce several proteins once it has reached the target region. Also, transporting mRNA, prevents the corresponding proteins from unwanted interactions on their way to the target site. Another benefit of the sub cellular localization of mRNA is facilitating production of proteins with high concentration which is beneficial for integration of proteins in macromolecular complexes.

There are three distinct mechanisms involved in the process of distribution of mRNAs within cells: localized protection from degradation, uniform diffusion coupled with local entrapment and directed transport. We are interested in the third mechanism. *RNA binding proteins (RBPs*) play an important role in the active transport of mRNA. RBPs bind to target mRNAs through cis-localization elements (or zipcodes), and make a complex that transports the mRNA. Furthermore, the structure of an RNA molecule plays key roles in determining the accessibility of these binding sites. Despite their importance in gene regulation, the binding sites for most RBPs involved in mRNA localization are not well characterized.

In this report, we develop a pipeline for predicting the localization patterns of mRNAs in Drosophila melanogaster embryos, and use interpretable machine learning model to identify RBPs involved in this process. We present the work of Ribeiro *et al.* [1] who propose LIME, an explanation technique which attempts to provide interpretable explanations for predictions of *any* classifier.

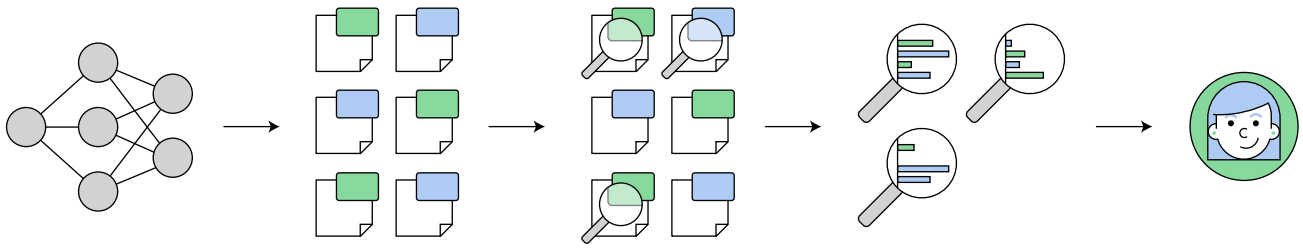
LIME approaches the problem by learning an *interpretable model* (called the *explainer*) in the vicinity (in feature space) of the prediction generated by the ML model. This report employs LIME to reproduce the results presented by the authors and to further evaluate the methodology on the given dataset.

II. Method Overview

In this section, we introduce the notion of *explanation* and formalize its definition by outlining the algorithm discussed in [1]. A high-level description of the LIME pipeline is given in Figure 1. LIME stands for Local Interpretable Model-Agnostic Explanations.

A. Overview of an Explainer

There are four important desired characteristics of a *Post-hoc explainer*. First and foremost, it must be *interpretable*, meaning that it should provide a qualitative understanding between the input and the output. Note that the notion of interpretability directly depends on the target audience, as one type of explainer could be more easily interpreted by machine learning experts than by laymen. Second, it should conform to the notion of *local fidelity*. This implies that we require an *explainer* to be locally faithful in order for it to be meaningful. Put differently, it must comply to how the model behaves in the vicinity of the instance being predicted. It should be noted that local fidelity does not imply global fidelity. Third, the *explainer* should be *model-agnostic*, *i.e.*, it should treat any ML model as a black box. This ensures that the tool will be able to explain both the contemporary state-of-the-art models as well as future, more complex classifiers. Finally, we want an *explainer* to provide a *global perspective* on how the model behaves. This enables us to evaluate the performance of a model and conduct comparisons between different models before real-world deployments.



(a) Model (b) Predictions (c) Pick step (d) Explanations (e) User

Fig. 1. LIME PIPELINE. A complex model (a) is used to predict labels from a given dataset (b). Then, an explainer is learned to choose representative instances (c) to explain. The explainer provides explanations (d) to the user, who takes the final decision (e).

B. An Algorithm for Explaining

More formally, an *explainer* is defined as a model $g \in \mathcal{G}$ where $\mathcal{G}$ is the set of potentially interpretable models such as linear models and decision trees. Given an original instance $x \in \mathbb{R}^d$ to be explained, we first convert it to its *interpretable representation*. An *interpretable representation* is any representation that is intuitive for humans, *i.e.*, it should conform to the notion of *decomposability*. LIME denotes this representation as a binary vector $\varphi(x) = \widehat{x} \in \{0,1\}^{\widehat{d}}$. In the case of text classification this denotes the presence or absence of a word while for image classification it represents presence or absence of a contiguous patch of pixels (a superpixel).

Though models like linear models and decision trees are considered interpretable, a large dimensional model can lead to the degradation in its *simulability*. Hence the authors define $\Omega(g)$ to be a measure of complexity of the explainer. This can be the depth of a tree for decision trees, or the number of nonzero weights for a linear model.

Now, let $f : \mathbb{R}^d \to \mathcal{C}$ be a classifier over a set of classes $\mathcal{C}$, to be explained. Further let $\pi_x(z)$ be a proximity metric between x and z that encapsulates the notion of locality around x. Given these definitions, we set the notion of *local fidelity* by defining a loss function $L(f, g, \pi_x)$ which represents the explainer's faithfulness to the original model in the locality of the x as defined by $\pi_x(z)$. To obtain an interpretable explanation, we minimize this loss function while also ensuring *simulability* by keeping $\Omega(g)$ low. This is given by

$$\xi(x) = \underset{g \in \mathcal{G}}{\arg\min}\ L(f, g, \pi_x) + \Omega(x). \tag{1}$$

To approximate L, we sample around $\widehat{x}$ to get a set of perturbed instances $\widehat{z} \in \{0,1\}^{\widehat{d}}$. The perturbed samples $\widehat{z}$ are then transformed to the original representation of the instance to get $z \in \mathbb{R}^d$. This is done by letting $\widehat{z}$ act as a binary mask over x. $f(z)$ now denotes the labels generated by the original classifier for the perturbed samples. Assuming we let $\pi_x(z) = \exp(-d(x,z)^2/\sigma^2)$ be an exponential kernel with variance σ^2 defined on some distance function d (*e.g.*, cosine distance for text), then the locality-aware loss L is given by

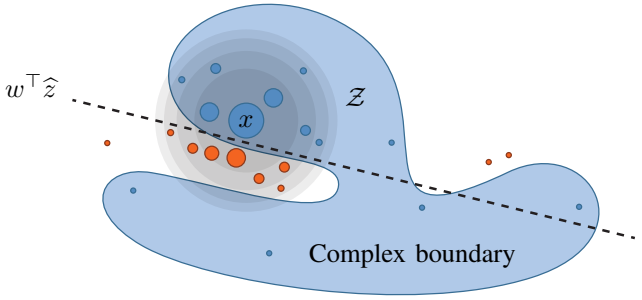


Fig. 2. INTUITION BEHIND LIME. Points are sampled around x and a new, simpler decision boundary is learned as an explanation.

$$L(f, g, \pi_x) = \sum_{z, \widehat{z} \in \mathcal{Z}} \pi_x(z)\big(f(z) - g(\widehat{z})\big)^2, \tag{2}$$

where $\mathcal{Z}$ is the perturbed samples set with associated labels. LIME currently defines g as a linear classifier, that is, $g(\widehat{z}) = w^\top \widehat{z}$ for some weight vector w. This definition can however be extended to include other explainers such as decision trees. Figure 2 illustrates the above procedure with a linear explainer.

As mentioned earlier, in the case of text classification the interpretable representation is denoted by the presence or absence of a word. Hence the complexity parameter $\Omega(x)$ is specified by a limit K on the number of words to be accounted in the computation. In the of case images the authors use *superpixels* as the interpretable representation. Using the same notion the complexity parameter $\Omega(x)$ now sets a limit on the number of superpixels to be considered during computation. This renders the Equation 1 intractable. Authors address this problem by introducing the K-Lasso algorithm, where $\Omega(x)$ is approximated by selecting K features using the L_1-regularization (Lasso) and then learning the weights via least squares procedure. This describes the formulation of an *explainer* $\xi(x) = (g \circ f)(x)$.

The second part of the algorithm consists of providing the user with a global representation of a model. This is achieved via judiciously selecting multiple instances as representation of the model to be presented to the user. Put briefly, a Submodular Pick (SP) algorithm is defined to choose a diverse,

representative set of explanations to assist in understanding how a model behaves globally. This essentially translates into a max-coverage problem of size B (which denoted the user budget) which can be well-approximated by a greedy algorithm, offering a constant-factor approximation guarantee of $1 - 1/e$ to the optimum. For further details of the SP algorithm the reader is suggested to refer to [1].

C. Algorithm Implementation

LIME is an open-source project available at [3]. Experiments are conducted by modifying the given framework in order to accommodate different classifiers and datasets. For text classification, the cosine distance for d is used and perturbations to documents are performed by randomly removing words. For image classification, the L^2 (Euclidean) distance is employed, and perturbed instances are created by segmenting them into contiguous superpixels and randomly removing patches using [4]. This process is depicted in Figure 3.

(a) Original image (b) Superpixel segments

Fig. 3. SUPERPIXEL ALGORITHM. Each of the $n = 30$ patches corresponds to an interpretable component for LIME.

TensorFlow [5] is employed to easily obtain complex deep convolutional neural network architectures. For all other classifiers, scikit-learn [6] modules are used.

III. LIME ON BREAST CANCER TUMOR DATA SET

We begin with evaluation of LIME to explain predictions made on medical data using the breast cancer dataset from scikit-learn, which a copy of UCI ML Breast Cancer Wisconsin (Diagnostic) Dataset [7]. Medical applications such as this serves as a major motivation for explanation tools like LIME. It is often necessary to inspect the reasons of a particular prediction generated by a ML model considering the sensitive nature of the application. We illustrate such a scenario with the suggested evaluation.

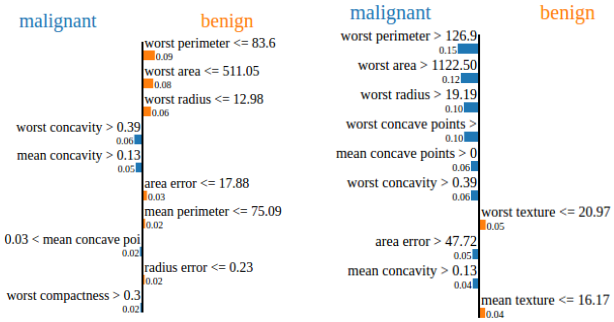
The features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. These describe characteristics of the cell nuclei present in the image scan. The dataset includes 569 instances and two classes: *malignant* and *benign*, with 212 and 357 instances, respectively. Ten core attributes are used as detailed in Table I. The mean, standard error, and "worst" or largest (mean of the three largest values) of these features are computed for each instance, amounting to a total of 30 features.

TABLE I
UCI BREAST CANCER DATASET CORE ATTRIBUTES

FEATURE	DESCRIPTION
Radius	Mean of distances from center to points on the perimeter
Texture	Standard deviation of grayscale values
Perimeter	Distance around the nuclear border
Area	Area of outlined nucleus
Smoothness	Local variation in radius lengths
Compactness	Perimeter^2 / area $-$ 1.0
Concavity	Severity of concave portions of the contour
Concave points	Number of concave portions of the contour
Symmetry	Symmetry of the outlined nucleus
Fractal dimension	Coastline approximation $-$ 1.0

Random Forests with default parameters and $n = 500$ trees is used to model the data. Random Forests are chosen due to their complex nature of decision logic. This is a good example of the types of models LIME will have to explain in the field. We also use a second classifier, a Gaussian Naive Bayes classifier, to model the data. The training and test sets are divided in a ratio of 80/20. For explaining individual test cases, predicted instances for both classifiers are obtained and an explainer is learned on them to analyze which features are being used to generate the prediction.

The results for breast cancer tumors show that the algorithm is capable of identifying relevant features that contribute to a prediction. Here, *malignant* and *benign* cases are taken individually and an explainer is learned on the predictions. The results are shown in Figure 4.



(a) *Benign* Prediction (b) *Malignant* Prediction

Fig. 4. EXPLAINING BENIGN AND MALIGNANT CASES.

For both the instances, worst concavity adds weight to a *malignant* prediction (> 0.39) but in the *benign* case, other features (worst perimeter, area, radius) add more weight to the categorization task.

One of the assumptions of using the explainer is that you need expert/background knowledge in order to interpret the results. When comparing the results with these in the literature [8], it was found that the results obtained are akin to implying that if used in a medical setting, providing explanations can increase the acceptance of automated systems.

IV. Predicting Localization Patterns of mRNAs

In this section, we describe the process to identify RBPs in predicting the localization patterns of mRNAs.

A. Data Set Description

We ran our predictor on 6 different data sets containing 5 classes - Table II. Out of these, one dataset was imbalanced which resulted in high accuracy but low precision/recall. The rest of datasets contains 300 feature vectors and were balanced for both classes. A feature vector (n1, n2, , nm) is made for each gene, where coordinates correspond to number of times each RBP binding site is present in paired/unpaired regions.

Labels 0 and 1 indicate whether the RNAs bind or not. To make the predictions, training and validation set were divided in 80/20 for each of the classes.

TABLE II
Classes in the dataset provided

Classes
Cortical Nuclei
Perinuclear around pole cell nuclei
RNA islands
Yolk plasm enrichment
Pole plasm

B. Prediction

Random Forest Classifier was used and the results obtained on each of the dataset are given in Table III. Also, the ROC curve for all are given in the Appendix. As seen in the table, even though the accuracy was good for imbalanced dataset, recall was just 29.5%. For this reason, we run LIME on only balanced datasets, so that there is no bias in explaining the predictions.

TABLE III
Accuracy, Precision and Recall

Classes	Accuracy	Precision	Recall
Unbalanced Cortical Nuclei	94.3	91.9	29.5
Cortical Nuclei	83.3	79.5	81.8
Perinuclear around pole cell nuclei	98.8	97.5	100
RNA islands	84.6	79.2	85.1
Yolk Plasm enrichment	79.2	74.9	85.8
Pole Plasm	80.12	75.8	86.28

C. Explaining the Prediction - Identifying RBPs

To identify RBP's, we run explainer on individual instances. For this report, we are focusing on Cortical Nuclei class only. One of the observations was that the key feature vectors which motivated the prediction had quite less weight (e-50) for this dataset unlike other datasets which had a positive weight in the range 0-10 for each of the key feature vectors.

Everytime the explainer was run, top features changed (due to negligible weight). In order to get meaningful results, we ran the loop 10 times for Cortical nuclei class on the true positive predictions and made a note of the features which were repeated as shown in Table IV, We identify these as RBPs essential for making the prediction in this process.

TABLE IV
Key RBPs identified for some of the True positive predictions on Cortical Nuclei Class

Name	RBP's identified
FBtr0075329-apiMel4.Group1	Rna15, ybx2-a, SNRPA
FBtr0075329-dm6.chr3L	SRSF2, SNRPA
FBtr0075329-droBip2.KB463974	SNRPA, SNRPA, SRSF1, SRSF2, SNRPA
FBtr0075329-droEug2.KB465257	SRSF2, SNRPA, ybx2-a
FBtr0075329-droKik2.KB459701	SNRPA , SNRPA, SNRPA
FBtr0075329-droMoj3.scaffold_6680	SRSF2, SRSF7, ELAVL2, SRSF2, ybx2-a
FBtr0075329-droPer1.super_29	SRSF7, SNRPA

The list of key feature vectors for all true positive cases which were selected by Lime algorithm have been attached in the appendix for all true positives of class Cortical Nuclei.

V. Future Work

Next steps would be to compare the key RBPs identified by this process with that of Lasso classifier and to see if we can find some interesting findings in that.

References

[1] M. T. Ribeiro, S. Singh, and C. Guestrin, ""Why should I trust you?": Explaining the predictions of any classifier," *CoRR*, vol. abs/1602.04938, 2016. [Online]. Available: http://arxiv.org/abs/1602.04938

[2] "Global analysis of mrna localization reveals a prominent role in organizing cellular architecture and function," vol. 131, pp. 174–187, 2007.

[3] M. T. Ribeiro, "Lime: Explaining the predictions of any machine learning classifier," https://github.com/marcotcr/lime, 2017.

[4] X. Ren and J. Malik, "Learning a classification model for segmentation," in *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2*, ser. ICCV '03. Washington, DC, USA: IEEE Computer Society, 2003, pp. 10–. [Online]. Available: http://dl.acm.org/citation.cfm?id=946247.946677

[5] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: http://tensorflow.org/

[6] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, and G. Varoquaux, "API design for machine learning software: experiences from the scikit-learn project," in *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, 2013, pp. 108–122.

[7] M. Lichman, "UCI machine learning repository," 2013. [Online]. Available: http://archive.ics.uci.edu/ml

[8] B. V. Aparna Narasimha and H. M. Kumar, "Significance of nuclear morphometry in benign and malignant breast aspirates," *International Journal of Applied Basic Medical Research*, 2013.

Appendix

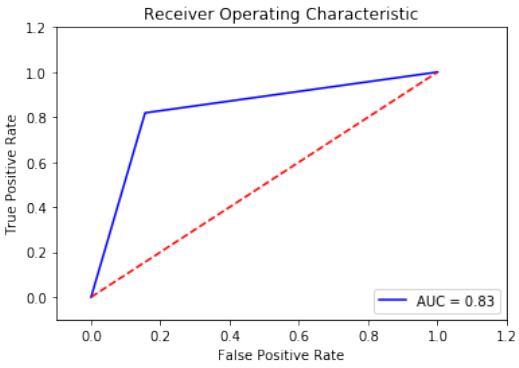


Fig. 5. ROC Curve For Cortical Nuclei Class

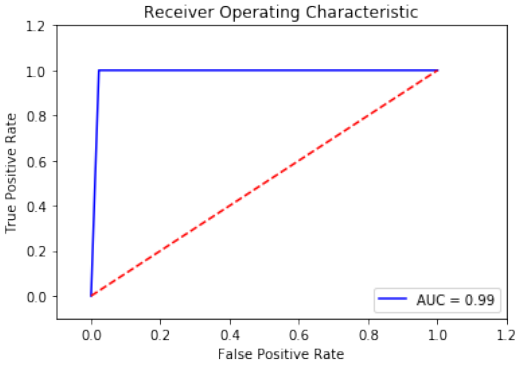


Fig. 6. ROC Curve For Perinuclear around pole cell nuclei Class

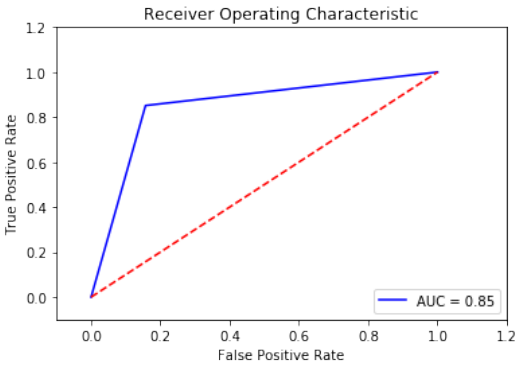


Fig. 7. ROC Curve For RNA Island Class

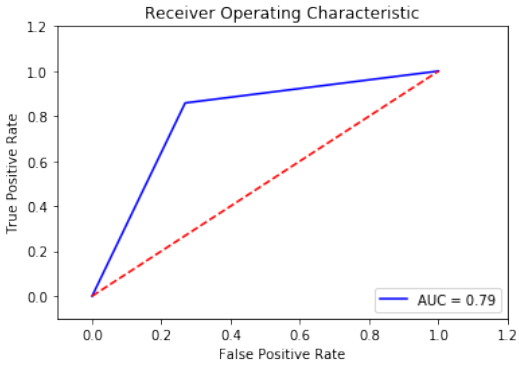


Fig. 8. ROC Curve For Yolk plasm enrichment Class

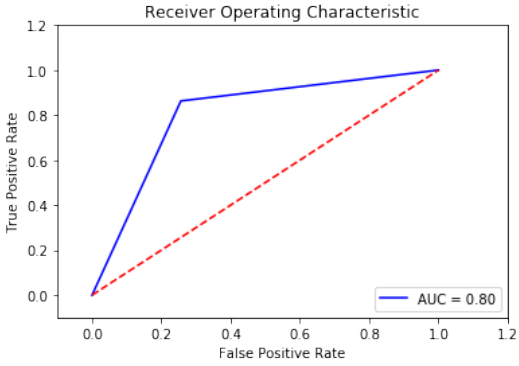


Fig. 9. ROC Curve For Pole plasm Class

ID: 0
Name: FBtr0075329-apiMel4.Group1
[('267__Rna15 <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 1
Name: FBtr0075329-dm6.chr3L
[('256__SRSF2 <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 5
Name: FBtr0075329-droBip2.KB463974
[('285__SNRPA <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0), ('289__SRSF1 <= 0.00', 0.0), ('279__SRSF2 <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0)]
ID: 8
Name: FBtr0075329-droEug2.KB465257
[('256__SRSF2 <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0)]
ID: 10
Name: FBtr0075329-droKik2.KB459701
[('285__SNRPA <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 12
Name: FBtr0075329-droMoj3.scaffold_6680
[('256__SRSF2 <= 0.00', 0.0), ('278__SRSF7 <= 0.00', 0.0), ('275__ELAVL2 <= 0.00', 0.0), ('279__SRSF2 <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0)]
ID: 13
Name: FBtr0075329-droPer1.super_29
[('278__SRSF7 <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 15
Name: FBtr0075329-droRho2.KB452318
[('266__Rna15 <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 16
Name: FBtr0075329-droSec1.super_0
[('263__IGF2BP1 <= 0.00', 0.0), ('279__SRSF2 <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 19
Name: FBtr0075329-droVir3.scaffold_13049
[('256__SRSF2 <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0)]
ID: 20
Name: FBtr0075329-droYak3.chr3L
[('267__Rna15 <= 0.00', 0.0), ('258__su(s) <= 0.00', 0.0), ('278__SRSF7 <= 0.00', 0.0), ('256__SRSF2 <= 0.00', 0.0)]
ID: 22
Name: FBtr0083212-droAna3.scaffold_13340
[('267__Rna15 <= 0.00', 0.0), ('278__SRSF7 <= 0.00', 0.0), ('275__ELAVL2 <= 0.00', 0.0)]
ID: 24
Name: FBtr0083212-droBip2.KB464369
[('278__SRSF7 <= 0.00', 0.0)]
ID: 26
Name: FBtr0083212-droEre2.scaffold_4770
[('267__Rna15 <= 0.00', 0.0), ('258__su(s) <= 0.00', 0.0), ('256__SRSF2 <= 0.00', 0.0)]
ID: 32
Name: FBtr0083212-droPer1.super_3
[('285__SNRPA <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 33
Name: FBtr0083212-droRho2.KB451728

[('249__ybx2-a <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0), ('274__NOVA2 <= 0.00', 0.0)]
ID: 36
Name: FBtr0083212-droSuz1.KI420265
[('249__ybx2-a <= 0.00', 0.0), ('275__ELAVL2 <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 39
Name: FBtr0083212-droYak3.chr3R
[('256__SRSF2 <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0)]
ID: 43
Name: FBtr0073301-droBia2.KB462838
[('284__SNRPA <= 0.00', 0.0), ('274__NOVA2 <= 0.00', 0.0)]
ID: 44
Name: FBtr0073301-droBip2.KB464401
[('284__SNRPA <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 47
Name: FBtr0073301-droEug2.KB465069
[]
ID: 55
Name: FBtr0073301-droSim1.chr3L
[('249__ybx2-a <= 0.00', 0.0), ('263__IGF2BP1 <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 57
Name: FBtr0073301-droVir3.scaffold_13049
[('266__Rna15 <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 60
Name: FBtr0073301-musDom2.KB855382
[('249__ybx2-a <= 0.00', 0.0), ('278__SRSF7 <= 0.00', 0.0)]
ID: 65
Name: FBtr0075053-droEre2.scaffold_4784
[('283__SNRPA <= 0.00', 0.0), ('278__SRSF7 <= 0.00', 0.0), ('279__SRSF2 <= 0.00', 0.0), ('274__NOVA2 <= 0.00', 0.0)]
ID: 72
Name: FBtr0075053-droPer1.super_9
[('256__SRSF2 <= 0.00', 0.0), ('263__IGF2BP1 <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 73
Name: FBtr0075053-droVir3.scaffold_13049
[('258__su(s) <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 74
Name: FBtr0075053-droWil2.CH964095
[]
ID: 76
Name: FBtr0075053-musDom2.KB860774
[('256__SRSF2 <= 0.00', 0.0)]
ID: 79
Name: FBtr0080638-droEle2.KB458429
[('285__SNRPA <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0), ('275__ELAVL2 <= 0.00', 0.0), ('289__SRSF1 <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0)]
ID: 80
Name: FBtr0080638-droEre2.scaffold_4929
[('267__Rna15 <= 0.00', 0.0), ('275__ELAVL2 <= 0.00', 0.0), ('279__SRSF2 <= 0.00', 0.0)]
ID: 83
Name: FBtr0080638-droKik2.KB459688
[]
ID: 86

Name: FBtr0080638-droPer1.super_5
[('275__ELAVL2 <= 0.00', 0.0), ('266__Rna15 <= 0.00', 0.0)]
ID: 100
Name: FBtr0083083-droEre2.scaffold_4770
[('275__ELAVL2 <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 103
Name: FBtr0083083-droGri2.scaffold_14906
[]
ID: 106
Name: FBtr0083083-droSim1.chr3R
[('283__SNRPA <= 0.00', 0.0)]
ID: 108
Name: FBtr0083083-droWil2.CH964232
[('267__Rna15 <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0)]
ID: 109
Name: FBtr0083083-droYak3.chr3R
[('256__SRSF2 <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0), ('275__ELAVL2 <=
0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 116
Name: FBtr0088676-droFic2.KB457333
[('275__ELAVL2 <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0), ('283__SNRPA <=
0.00', 0.0)]
ID: 118
Name: FBtr0088676-droKik2.KB459872
[('285__SNRPA <= 0.00', 0.0), ('263__IGF2BP1 <= 0.00', 0.0), ('284__SNRPA <=
0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 122
Name: FBtr0088676-droSuz1.KI420922
[('249__ybx2-a <= 0.00', 0.0)]
ID: 124
Name: FBtr0088676-droWil2.CH963719
[('249__ybx2-a <= 0.00', 0.0), ('278__SRSF7 <= 0.00', 0.0), ('279__SRSF2 <=
0.00', 0.0), ('274__NOVA2 <= 0.00', 0.0)]
ID: 127
Name: FBtr0080389-dm6.chr2L
[('263__IGF2BP1 <= 0.00', 0.0), ('266__Rna15 <= 0.00', 0.0)]
ID: 135
Name: FBtr0080389-droFic2.KB457472
[('256__SRSF2 <= 0.00', 0.0), ('285__SNRPA <= 0.00', 0.0), ('284__SNRPA <=
0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 140
Name: FBtr0080389-droPer1.super_1
[('267__Rna15 <= 0.00', 0.0), ('258__su(s) <= 0.00', 0.0), ('256__SRSF2 <=
0.00', 0.0)]
ID: 141
Name: FBtr0080389-droPse3.chr4_CH379060_3_random
[('256__SRSF2 <= 0.00', 0.0), ('274__NOVA2 <= 0.00', 0.0), ('278__SRSF7 <=
0.00', 0.0), ('266__Rna15 <= 0.00', 0.0)]
ID: 144
Name: FBtr0080389-droSim1.chr2L
[('283__SNRPA <= 0.00', 0.0)]
ID: 146
Name: FBtr0080389-droVir3.scaffold_12963
[('258__su(s) <= 0.00', 0.0), ('266__Rna15 <= 0.00', 0.0), ('249__ybx2-a <=
0.00', 0.0)]
ID: 150
Name: FBtr0300090-droSec1.super_20
[('256__SRSF2 <= 0.00', 0.0), ('263__IGF2BP1 <= 0.00', 0.0), ('275__ELAVL2
<= 0.00', 0.0), ('279__SRSF2 <= 0.00', 0.0)]

ID: 153
Name: FBtr0075167-droAlb1.JH856983
[('258__su(s) <= 0.00', 0.0), ('263__IGF2BP1 <= 0.00', 0.0), ('279__SRSF2 <= 0.00', 0.0)]
ID: 154
Name: FBtr0075167-droAna3.scaffold_13337
[('256__SRSF2 <= 0.00', 0.0), ('274__NOVA2 <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0)]
ID: 155
Name: FBtr0075167-droBia2.KB462646
[('285__SNRPA <= 0.00', 0.0), ('289__SRSF1 <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 162
Name: FBtr0075167-droKik2.KB459821
[('256__SRSF2 <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 163
Name: FBtr0075167-droMir2.chrXR
[('285__SNRPA <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0), ('289__SRSF1 <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0), ('266__Rna15 <= 0.00', 0.0)]
ID: 165
Name: FBtr0075167-droPer1.super_20
[('274__NOVA2 <= 0.00', 0.0)]
ID: 166
Name: FBtr0075167-droPse3.chrXR_CH379069_3_random
[('256__SRSF2 <= 0.00', 0.0), ('279__SRSF2 <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0)]
ID: 172
Name: FBtr0075167-droYak3.chr3L
[('256__SRSF2 <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0), ('274__NOVA2 <= 0.00', 0.0)]
ID: 173
Name: FBtr0075218-dm6.chr3L
[('256__SRSF2 <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 175
Name: FBtr0075218-droAna3.scaffold_13337
[('256__SRSF2 <= 0.00', 0.0), ('274__NOVA2 <= 0.00', 0.0), ('279__SRSF2 <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 177
Name: FBtr0075218-droBip2.KB464219
[('256__SRSF2 <= 0.00', 0.0), ('279__SRSF2 <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 179
Name: FBtr0075218-droEre2.scaffold_4784
[('278__SRSF7 <= 0.00', 0.0), ('279__SRSF2 <= 0.00', 0.0), ('266__Rna15 <= 0.00', 0.0)]
ID: 185
Name: FBtr0075218-droMoj3.scaffold_6680
[('267__Rna15 <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0), ('256__SRSF2 <= 0.00', 0.0)]
ID: 191
Name: FBtr0075218-droSuz1.KI420009
[('278__SRSF7 <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 193
Name: FBtr0075218-droYak3.chr3L
[('285__SNRPA <= 0.00', 0.0), ('278__SRSF7 <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 195

Name: FBtr0077832-droAna3.scaffold_12916
[('258__su(s) <= 0.00', 0.0), ('278__SRSF7 <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 196
Name: FBtr0077832-droBia2.KB462402
[('258__su(s) <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0), ('278__SRSF7 <= 0.00', 0.0), ('275__ELAVL2 <= 0.00', 0.0), ('256__SRSF2 <= 0.00', 0.0)]
ID: 197
Name: FBtr0077832-droBip2.KB464241
[('285__SNRPA <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 202
Name: FBtr0077832-droGri2.scaffold_15252
[('267__Rna15 <= 0.00', 0.0), ('256__SRSF2 <= 0.00', 0.0), ('279__SRSF2 <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0)]
ID: 204
Name: FBtr0077832-droMir2.chr4
[('258__su(s) <= 0.00', 0.0), ('256__SRSF2 <= 0.00', 0.0), ('278__SRSF7 <= 0.00', 0.0), ('275__ELAVL2 <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0)]
ID: 205
Name: FBtr0077832-droMoj3.scaffold_6500
[('256__SRSF2 <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0), ('285__SNRPA <= 0.00', 0.0)]
ID: 206
Name: FBtr0077832-droPer1.super_1
[('263__IGF2BP1 <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0)]
ID: 207
Name: FBtr0077832-droPse3.chr4_CH379060_3_random
[('256__SRSF2 <= 0.00', 0.0), ('263__IGF2BP1 <= 0.00', 0.0), ('279__SRSF2 <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0)]
ID: 209
Name: FBtr0077832-droSec1.super_121
[('274__NOVA2 <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 210
Name: FBtr0077832-droSim1.chr2L
[('249__ybx2-a <= 0.00', 0.0)]
ID: 212
Name: FBtr0077832-droVir3.scaffold_12963
[('256__SRSF2 <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0), ('279__SRSF2 <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 214
Name: FBtr0077832-droYak3.chr2L
[('249__ybx2-a <= 0.00', 0.0)]
ID: 216
Name: FBtr0343860-dm6.chr2L
[('267__Rna15 <= 0.00', 0.0), ('256__SRSF2 <= 0.00', 0.0), ('263__IGF2BP1 <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 218
Name: FBtr0343860-droAna3.scaffold_12916
[('256__SRSF2 <= 0.00', 0.0), ('279__SRSF2 <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0)]
ID: 219
Name: FBtr0343860-droBia2.KB462819
[('258__su(s) <= 0.00', 0.0), ('263__IGF2BP1 <= 0.00', 0.0), ('256__SRSF2 <= 0.00', 0.0), ('279__SRSF2 <= 0.00', 0.0), ('266__Rna15 <= 0.00', 0.0), ('267__Rna15 <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0), ('278__SRSF7 <= 0.00', 0.0), ('274__NOVA2 <= 0.00', 0.0)]
ID: 220
Name: FBtr0343860-droBip2.KB464214

[('249__ybx2-a <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 222
Name: FBtr0343860-droEre2.scaffold_4845
[('258__su(s) <= 0.00', 0.0), ('278__SRSF7 <= 0.00', 0.0), ('266__Rna15 <= 0.00', 0.0)]
ID: 224
Name: FBtr0343860-droFic2.KB457391
[('267__Rna15 <= 0.00', 0.0), ('258__su(s) <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 231
Name: FBtr0343860-droSec1.super_7
[('256__SRSF2 <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 232
Name: FBtr0343860-droSim1.chr2L
[('266__Rna15 <= 0.00', 0.0), ('275__ELAVL2 <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 238
Name: FBtr0086180-anoGam1.chr2L
[('256__SRSF2 <= 0.00', 0.0), ('278__SRSF7 <= 0.00', 0.0), ('285__SNRPA <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 244
Name: FBtr0086180-droEre2.scaffold_4929
[('285__SNRPA <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0), ('266__Rna15 <= 0.00', 0.0)]
ID: 245
Name: FBtr0086180-droEug2.KB465221
[('267__Rna15 <= 0.00', 0.0), ('256__SRSF2 <= 0.00', 0.0), ('275__ELAVL2 <= 0.00', 0.0)]
ID: 246
Name: FBtr0086180-droFic2.KB457417
[('258__su(s) <= 0.00', 0.0), ('256__SRSF2 <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 256
Name: FBtr0086180-droSuz1.KI420922
[('284__SNRPA <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 257
Name: FBtr0086180-droVir3.scaffold_10324
[('258__su(s) <= 0.00', 0.0), ('279__SRSF2 <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 262
Name: FBtr0089395-droAlb1.JH862323
[('285__SNRPA <= 0.00', 0.0), ('274__NOVA2 <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0), ('289__SRSF1 <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 265
Name: FBtr0089395-droBip2.KB464214
[('267__Rna15 <= 0.00', 0.0), ('258__su(s) <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 270
Name: FBtr0089395-droGri2.scaffold_15126
[('258__su(s) <= 0.00', 0.0), ('275__ELAVL2 <= 0.00', 0.0), ('274__NOVA2 <= 0.00', 0.0)]
ID: 277
Name: FBtr0089395-droSec1.super_7
[('258__su(s) <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0)]
ID: 289
Name: FBtr0086301-droBia2.KB462718
[('266__Rna15 <= 0.00', 0.0)]

ID: 291
Name: FBtr0086301-droEle2.KB458561
[('256__SRSF2 <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0)]
ID: 296
Name: FBtr0086301-droKik2.KB459686
[('256__SRSF2 <= 0.00', 0.0), ('263__IGF2BP1 <= 0.00', 0.0), ('275__ELAVL2 <= 0.00', 0.0)]
ID: 297
Name: FBtr0086301-droMir2.chr3
[('267__Rna15 <= 0.00', 0.0), ('275__ELAVL2 <= 0.00', 0.0)]
ID: 298
Name: FBtr0086301-droMoj3.scaffold_6496
[('258__su(s) <= 0.00', 0.0), ('256__SRSF2 <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0), ('285__SNRPA <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0)]
ID: 301
Name: FBtr0086301-droRho2.KB451266
[('284__SNRPA <= 0.00', 0.0), ('266__Rna15 <= 0.00', 0.0)]
ID: 302
Name: FBtr0086301-droSec1.super_1
[('256__SRSF2 <= 0.00', 0.0), ('274__NOVA2 <= 0.00', 0.0)]
ID: 304
Name: FBtr0086301-droVir3.scaffold_12875
[('258__su(s) <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0), ('256__SRSF2 <= 0.00', 0.0)]
ID: 307
Name: FBtr0086301-musDom2.KB857006
[('267__Rna15 <= 0.00', 0.0), ('278__SRSF7 <= 0.00', 0.0), ('275__ELAVL2 <= 0.00', 0.0), ('279__SRSF2 <= 0.00', 0.0)]
ID: 312
Name: FBtr0087939-droBia2.KB462460
[('267__Rna15 <= 0.00', 0.0), ('256__SRSF2 <= 0.00', 0.0), ('274__NOVA2 <= 0.00', 0.0)]
ID: 314
Name: FBtr0087939-droFic2.KB457278
[('258__su(s) <= 0.00', 0.0), ('263__IGF2BP1 <= 0.00', 0.0), ('266__Rna15 <= 0.00', 0.0), ('274__NOVA2 <= 0.00', 0.0)]
ID: 320
Name: FBtr0344852-dm6.chr2L
[('284__SNRPA <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 323
Name: FBtr0344852-droBip2.KB464229
[('258__su(s) <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0), ('256__SRSF2 <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0)]
ID: 330
Name: FBtr0344852-droMoj3.scaffold_6500
[('256__SRSF2 <= 0.00', 0.0), ('274__NOVA2 <= 0.00', 0.0)]
ID: 331
Name: FBtr0344852-droSec1.super_7
[('283__SNRPA <= 0.00', 0.0), ('266__Rna15 <= 0.00', 0.0)]
ID: 332
Name: FBtr0344852-droSim1.chr2L
[('267__Rna15 <= 0.00', 0.0)]
ID: 340
Name: FBtr0079168-droEle2.KB458409
[('249__ybx2-a <= 0.00', 0.0), ('279__SRSF2 <= 0.00', 0.0), ('266__Rna15 <= 0.00', 0.0)]
ID: 341
Name: FBtr0079168-droEre2.scaffold_4929

[]
ID: 342
Name: FBtr0079168-droEug2.AFPQ02004873
[('249__ybx2-a <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 343
Name: FBtr0079168-droFic2.KB457397
[('256__SRSF2 <= 0.00', 0.0), ('263__IGF2BP1 <= 0.00', 0.0), ('275__ELAVL2 <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0)]
ID: 347
Name: FBtr0079168-droPer1.super_10
[('267__Rna15 <= 0.00', 0.0), ('256__SRSF2 <= 0.00', 0.0), ('263__IGF2BP1 <= 0.00', 0.0)]
ID: 350
Name: FBtr0079168-droSim1.chr2L
[('278__SRSF7 <= 0.00', 0.0)]
ID: 351
Name: FBtr0079168-droSuz1.KI419149
[('256__SRSF2 <= 0.00', 0.0)]
ID: 352
Name: FBtr0079168-droVir3.scaffold_12963
[('258__su(s) <= 0.00', 0.0), ('256__SRSF2 <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0), ('267__Rna15 <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0), ('278__SRSF7 <= 0.00', 0.0)]
ID: 353
Name: FBtr0079168-droWil2.CH963857
[('279__SRSF2 <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0), ('289__SRSF1 <= 0.00', 0.0), ('285__SNRPA <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0), ('274__NOVA2 <= 0.00', 0.0)]
ID: 356
Name: FBtr0300085-dm6.chrX
[('256__SRSF2 <= 0.00', 0.0), ('278__SRSF7 <= 0.00', 0.0)]
ID: 357
Name: FBtr0083417-anoGam1.chr2L
[('249__ybx2-a <= 0.00', 0.0), ('275__ELAVL2 <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 361
Name: FBtr0083417-droAna3.scaffold_13340
[('263__IGF2BP1 <= 0.00', 0.0), ('266__Rna15 <= 0.00', 0.0)]
ID: 362
Name: FBtr0083417-droBia2.KB462558
[('285__SNRPA <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0), ('289__SRSF1 <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 371
Name: FBtr0083417-droSec1.super_12
[('249__ybx2-a <= 0.00', 0.0), ('256__SRSF2 <= 0.00', 0.0)]
ID: 374
Name: FBtr0070233-droAna3.scaffold_13117
[('249__ybx2-a <= 0.00', 0.0), ('266__Rna15 <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 375
Name: FBtr0070233-droBia2.KB462463
[('279__SRSF2 <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]
ID: 377
Name: FBtr0070233-droEle2.KB458387
[('256__SRSF2 <= 0.00', 0.0), ('266__Rna15 <= 0.00', 0.0), ('275__ELAVL2 <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0)]
ID: 381

Name: FBtr0070233-droMir2.chrXR
[('285__SNRPA <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0), ('279__SRSF2 <= 0.00', 0.0), ('266__Rna15 <= 0.00', 0.0)]
ID: 382
Name: FBtr0070233-droPer1.super_11
[('258__su(s) <= 0.00', 0.0), ('275__ELAVL2 <= 0.00', 0.0), ('279__SRSF2 <= 0.00', 0.0), ('256__SRSF2 <= 0.00', 0.0)]
ID: 384
Name: FBtr0070233-droRho2.KB451367
[('283__SNRPA <= 0.00', 0.0), ('289__SRSF1 <= 0.00', 0.0), ('285__SNRPA <= 0.00', 0.0), ('278__SRSF7 <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0), ('274__NOVA2 <= 0.00', 0.0)]
ID: 389
Name: FBtr0334286-dm6.chr3R
[('275__ELAVL2 <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0)]
ID: 391
Name: FBtr0334286-droAna3.scaffold_13340
[('266__Rna15 <= 0.00', 0.0)]
ID: 392
Name: FBtr0334286-droBia2.KB462814
[('267__Rna15 <= 0.00', 0.0), ('258__su(s) <= 0.00', 0.0), ('256__SRSF2 <= 0.00', 0.0), ('275__ELAVL2 <= 0.00', 0.0), ('274__NOVA2 <= 0.00', 0.0)]
ID: 393
Name: FBtr0334286-droBip2.KB464355
[('267__Rna15 <= 0.00', 0.0), ('256__SRSF2 <= 0.00', 0.0), ('263__IGF2BP1 <= 0.00', 0.0), ('275__ELAVL2 <= 0.00', 0.0)]
ID: 394
Name: FBtr0334286-droEle2.KB457684
[('256__SRSF2 <= 0.00', 0.0), ('274__NOVA2 <= 0.00', 0.0), ('249__ybx2-a <= 0.00', 0.0)]
ID: 396
Name: FBtr0334286-droEug2.KB464661
[('258__su(s) <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0), ('275__ELAVL2 <= 0.00', 0.0), ('284__SNRPA <= 0.00', 0.0), ('266__Rna15 <= 0.00', 0.0)]
ID: 397
Name: FBtr0334286-droFic2.KB457554
[('279__SRSF2 <= 0.00', 0.0)]
ID: 401
Name: FBtr0334286-droMoj3.scaffold_6540
[('249__ybx2-a <= 0.00', 0.0), ('283__SNRPA <= 0.00', 0.0)]