# Teams recognized for developing top performing models to predict transcription factoring binding site across tissue types in a crowdsourced competition.

October 22, 2016

Transcription factors (TFs) are regulatory proteins that bind specific sequence motifs in the genome to activate or repress transcription of target genes. TFs play central roles in controlling biological processes and are often mis-regulated in disease. Identifying where TFs bind the genome is a fundamental, but elusive challenge. Improvements in next generation sequencing technologies have provided researchers a much deeper exploration of the binding patterns of TFs, but how to properly leverage these data to predict where a TF will bind across tissue and cell types remains an open question.

The ENCODE-DREAM in vivo Transcription Factor Binding Site Prediction Challenge was set up by a collaborative group of researchers to identify the best performing model for predicting positional in vivo TF binding sites. The Challenge presented several large sequencing-based datasets to the research community and international teams worked to address the challenge. Teams submitted their predictions and the challenge organizing team assessed method performance to produce an unbiased benchmarking of prediction models. The challenge is being run in two phases, with the first phase ending September 30, 2016 and the second phase ending January 11, 2017.

In total, 34 independent, international teams participated in the first phase of the challenge and two teams were named the top-performers. Team autosome.ru, from Russia, was the first place performer and J-TEAM, from Germany, was the second place performer. Team members are listed below. Both teams were invited to give a presentation of their methods at the RECOMB/ISCB Conference on Regulatory & Systems Genomics with DREAM Challenge & Cytoscape Workshop (https://www.iscb.org/recomb-regsysgen2016). The conference will be held on November 6-9, 2016 in Phoenix, Arizona, USA.

The top two teams were comprised of Bioinformatics specialists with expertise in regulatory genomics. Both teams built on their experience in TF binding site prediction and leveraged the challenge data and challenge framework to evaluate if their models held up to a novel dataset.

“The ENCODE-DREAM Challenge perfectly matches our research interests,” said autosome.ru team leader Ivan Kulakovskiy. “In the past, our team has developed several methods for analysis of transcription factor binding. We have got interested in DREAM challenges in 2013 when DREAM5 paper on transcription factor motifs was published, and we realized that our tool ChIPMunk was used as a baseline in benchmark on ChIP-Seq data.” The J-Team also expressed their excitement that the challenge closely matched their research expertise in de novo motif discovery and TF binding site prediction.

J-TEAM was able to leverage their past experience in DREAM challenges, in particular the DREAM5 TF-DNA Motif Recognition Challenge. To solve the ENCODE-DREAM challenge, J-TEAM developed a method termed Dimont. “Actually, an early prototype of the motif discovery approach Dimont has been developed for the DREAM5 challenge, so our approach directly profited from the results of this earlier challenge,” said J-TEAM leader Jan Grau.

While sequencing data are widely available, such as RNAseq and ChIPseq, some techniques are less common, such as the DNAseq provided in the challenge. Teams found these data useful in unexpected ways. “Even though we have been working with ENCODE data for several years, in this Challenge we have gained a lot of practical experience particularly in the analysis of deeply sequenced DNase tracks and in integrating several types of basic genomic data. Some ideas that seemed promising for the performance enhancement turned out to be worthless; such failures were almost as useful as our successful attempts,” said Ivan.

The DREAM Challenges are founded on the principles of building research communities and making data open to researchers around the world. Identification of top performing methods and benchmarking method performance are also key elements to the Challenges. While the challenges are academically rigorous, they are fun for participants. “In general, we think that challenges like this one are a unique opportunity to test your own methods and compare them with those of colleagues in an unbiased and fair manner. And, of course, its the fun and excitement of a true competition with unknown outcome, even after you pressed the ‘submit’ button,” said Jan.

The DREAM Challenges are team efforts and both teams wanted to emphasize the fact that each and every member of the team were major contributors to their solution. Well done automsome.ru and J-TEAM.

The second phase of the ENCODE-DREAM Challenge is open until January 11, 2017. For more details, see https://www.synapse.org/ENCODE

Top-performing Teams

**automosome.ru**

- Andrey Lando, Moscow Institute of Physics and Technology, *Dolgoprudny, Russia*
- Ilya Vorontsov, Vavilov Institute of General Genetics, Russian Academy of Sciences, *Moscow, Russia*
- Valentina Boeva, Institut Cochin, *Paris, France*
- Grigory Sapunov, Intento, *https://inten.to*
- Irina Eliseeva, Institute of Protein Research, Russian Academy of Sciences, *Pushchino, Russia*
- Vsevolod Makeev, Vavilov Institute of General Genetics, Russian Academy of Sciences, *Moscow, Russia*
- Ivan Kulakovskiy, Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, *Moscow, Russia*

**J-TEAM**

- Jens Keilwagen, Julius Kühn-Institut (JKI) - Federal Research Centre for Cultivated Plants, Quedlinburg, Germany
- Stefan Posch, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany
- Jan Grau, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany