Successful application of novel plant ChIP-seq technology results in high quality sequencing data from *Arabidopsis thaliana* seedling cells

Miklós Laczik, Céline Sabatel, Mareike Hohenstatt, Jan Hendrickx, Dominique Poncelet

Introduction

In the recent years massively parallel sequencing (also known as Next Generation Sequencing or NGS) has become the dominant method for sequencing, and the decreasing costs and increasing yields transformed biosciences. Due to greater accessibility and higher throughput, NGS is also highly utilized in epigenetic or epigenomic studies, most widely in ChIP-seq. NGS is the detection method of choice for ChIP, which gives an accurate qualitative and quantitative insight into protein-DNA interactions, such as binding events and sites of transcription factors or chromatin modifications including methylation and acetylation.

In addition to the widespread use of ChIP-seq and other NGS-related techniques, another major trend is the increasing use of limited amounts of starting material (e.g. due to ancient fossil samples, forensic samples, or limitations on amounts of material generated cells). In some instances, samples such as may not yield enough material for a regular ChIP-seq as in the case of embryonic tissues, the few cell layers of a specific zone in the apical meristem, etc. In other studies the researcher might be simply interested in the epigenetic events in a single cell.

Plant researchers also have a need for ChIP-seq from limited amounts of material, but their needs are rarely met. There are not many commercial solutions available that are dedicated to plant research, and usually these are not designed for low cell number experiments.

In this study, we show a successful and reliable ChIP-seq technology using limited amounts of plant material, with specific reagents and protocols, supported by strict bioinformatic QC analysis.

Methods

Arabidopsis thaliana is one of the most studied model organism in the plant research community, having given valuable insights into plant genetics, epigenetics, development and evolution. Therefore we chose *Arabidopsis thaliana* (ecotype Columbia) seedlings for our ChIP-seq experiment. We performed two series of biological replicates, with different amounts of starting materials, and as a reference we used samples processed with the Illumina TruSeq kit, and we also compared our results to public dataset of ChIP-seq on A. thaliana seedlings. ChIP assays were performed using the Diagenode Plant ChIP kit (Cat. No. C01010150) on 0.25 g freshly weighed plant tissue with the antibody against the histone mark H3K4me3 (Cat. No. C15410003). Bioruptor Pico (Cat. No. B0160001), thermo-controlled sonication device, which was used for the shearing of the chromatin. The immunoprecipitated DNA was used for library preparation using the MicroPlex Library Preparation kit (Cat. No. C05010010) on different amounts of DNA (1 ng, 500 pg and 100 pg) and using the Illumina TruSeq kit on 5 ng of DNA. Sequencing was done on Illumina HiSeq 2000. The overview of the experimental design is in Table 1.

We use our proprietary bioinformatics pipeline to analyse the data and perform a strict QC. Some of the QC methods are derived from the ENCODE/modENCODE projects, which are considered gold standards for ChIP-seq. See the concepts of FRIP, NSC, RSC in the paper detailing the ENCODE/modENCODE guidelines (Landt et al: ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. 2012 Sep;22(9):1813-31.) The overlap criteria are also defined by the ENCODE/modENCODE criteria, where at least 80% of the top 40% of the peaks of two replicates should have a match in the replicate dataset.