(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 April 2019 (25.04.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/076768 A1

(51) International Patent Classification:
C12Q 1/6806 (2018.01) *C12Q 1/6804* (2018.01)

(21) International Application Number:
PCT/EP2018/077934

(22) International Filing Date:
12 October 2018 (12.10.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
1751281-5 16 October 2017 (16.10.2017) SE

(71) Applicants: **TERVISETEHNOLOOGIATE ARENDUSKESKUS AS** [EE/EE]; Tiigi 61b, 50410 TARTU (EE). **TARTU ÜLIKOOL** [EE/EE]; Ülikooli 18, 50090 TARTU (EE).

(72) Inventors: **ORG, Tõnis**; Väike-Tähe 16-4, 51010 TARTU (EE). **KURG, Ants**; Liivakuru 3, 50304 TARTU (EE). **SALUMETS, Andres**; Hurda 31-1, 51005 TARTU (EE).

(74) Agent: **AROS PATENT AB**; Box 1544, 751 45 UPPSALA (SE).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
— *of inventorship (Rule 4.17(iv))*

Published:
— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) **Title:** METHOD AND KIT FOR DNA LIBRARY PREPARATION

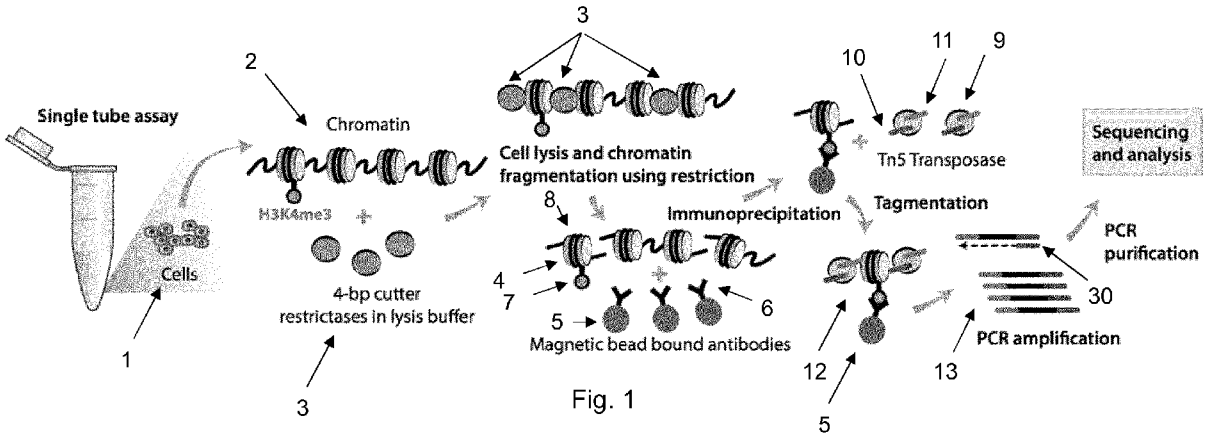


Fig. 1

(57) **Abstract:** A library of DNA molecules (13) is prepared from a chromatin sample (2) by enzymatic digestion of the chromatin sample (2). The resulting chromatin fragments (4) are immobilized onto a support (5) and tagged in a tagmentation process using a transposase (9). The DNA molecules (8) of the tagged chromatin fragments (12) immobilized onto the support (5) are amplified using forward and reverse amplification primers (20, 30) to form the library of amplified DNA molecules (13). The library preparation requires only very minute starting materials and is thereby useful in applications with limited starting material, such as highly pure cell populations or early developmental stages of mammalian preimplantation embryo.

WO 2019/076768 A1

METHOD AND KIT FOR DNA LIBRARY PREPARATION

TECHNICAL FIELD

The present embodiments generally relate to library preparation, and in particular to a method and a kit for library preparation from chromatin samples.

BACKGROUND

The development and functioning of an organism is determined by how the genetic information, encoded in its genomic deoxyribonucleic acid (DNA) sequence, is utilized by individual cells. In eukaryotic cells, DNA is packed into chromatin using proteins, of which histones are most abundant. Each cell type has its specific chromatin structure, which is dynamic and can be remodeled in response to extracellular signals to regulate gene expression, DNA repair and cell division. Our understanding of these chromatin-related processes has improved vastly over the past years thanks to the advances in DNA sequencing technologies.

Chromatin immunoprecipitation (ChIP) has been the method of choice to study the location of DNA bound proteins for years. Coupling ChIP with deep sequencing (ChIP-seq) has enabled determination of the localization of chromatin bound proteins at a genome-wide level. For example, it has helped identifying that different histone post-translational modifications are associated with different genomic features and transcriptional states, thereby contributing to explanation of how cell type specific gene regulation is achieved.

One of the limitations of the ChIP-seq methods is that they usually require large number of cells. In a typical experiment several million cells are used, which can be a limiting factor when working with samples where only limited numbers of cells are available, such as highly purified rare cell populations or early developmental stages. Recent advances in technologies have made it possible to develop more sensitive ChIP-seq methods. However, many of these methods are complex, laborious, require specific apparatus or are not sensitive enough.

A typical ChIP-seq experiment consists of several experimental steps to produce a library that can be sequenced using massively parallel sequencing. These steps include fixing, cell lysis, chromatin fragmentation, immunoprecipitation, decrosslinking, DNA purification, and sequencing library preparation followed by sequencing. There are many issues that arise when working with low numbers of cells, of which loss of material is the most prominent. In the published protocols, the reduction of material loss

has been achieved through the use of different carriers that mimic more material or with indexing first and then pooling the samples to obtain more material for subsequent steps.

There are many instances where more sensitive ChIP-seq methods could be useful but the most obvious is studying early development, like mammalian embryo preimplantation development, as the number of available cells in these study samples is especially low. The cellular changes that take place during development are remarkable both at the molecular and phenotypic level. Understanding the molecular basis for differentiation it is not only fascinating but it also holds great promise to advance reproductive medicine and the generation of desired cell types for regenerative medicine. Recently, the first histone modification landscapes of early mouse development were reported. In addition to using more sensitive ChIP assays a large number of embryos were still needed to be pooled in order to obtain sufficient number of cells. Therefore, although substantial advancements have been made both in sensitivity and simplicity of the ChIP-seq methods, there is still room for improvement. This is in particular important in the context of studying human embryonic development as many legal, ethical and technical issues come to play.

Accordingly, there is a need for a ChIP-method that can be used in applications with limited number of cells that would be sensitive but also robust so it could be used without the need of special equipment.

WO 2011/096926 relates to a chromatin immunoprecipitation-high throughput sequencing technique that allows creation of chromatin maps from limited biological sample sizes. The document, though, mentions that the biological sample has up to 50,000 cells, which is still regarded as a very large sample size in the context of studying human embryonic development.

SUMMARY

It is a general objective to provide a method and a kit for library preparation for chromatin samples.

It is a particular objective to provide such library preparation suitable for applications with limited sample material.

These and other objectives are met by embodiments as disclosed herein.

The invention is defined in the independent claims. Further embodiments of the invention are defined in the dependent claims.

An aspect of the embodiments relates to a method of library preparation. The method comprises enzymatically digesting a chromatin sample using at least one restriction enzyme to form a plurality of chromatin fragments. The method also comprises immobilizing at least a portion of the plurality of chromatin fragments onto at least one solid support comprising a respective affinity molecule having affinity for a polypeptide bound to DNA molecules in the at least a portion of the plurality of chromatin fragments. The method further comprises tagging the at least a portion of the plurality of chromatin fragments immobilized onto the at least one solid support in a tagmentation process using a transposase and at least one tagging adapter to form tagged chromatin fragments immobilized onto the at least one solid support. The method additionally comprises amplifying DNA molecules of the tagged chromatin fragments in presence of a forward amplification primer and a reverse amplification primer to form a library of amplified DNA molecules. The forward amplification primer comprises, from a 5' end to a 3' end, a first common sequence and a first sequence corresponding to at least a sequence portion of a tagging adapter of the at least one tagging adapter. The reverse amplification primer comprises, from a 5' end to a 3' end, a second common sequence and a second sequence corresponding to at least a sequence portion of a tagging adapter of the at least one tagging adapter.

Another aspect of the embodiments relates to a kit for library preparation. The kit comprises at least one restriction enzyme configured to enzymatically digest a chromatin sample to form a plurality of chromatin fragments. The kit also comprises at least one solid support comprising a respective affinity molecule having affinity for a polypeptide bound to DNA molecules in the at least a portion of the plurality of chromatin fragments. The kit further comprises a transposase and at least one tagging adapter. The transposase is configured to tag the at least a portion of the plurality of chromatin fragments immobilized onto the at least one solid support in a tagmentation process to form tagged chromatin fragments immobilized onto the at least one solid support. The kit also comprises a forward amplification primer comprising, from a 5' end to a 3' end, a first common sequence and a first sequence corresponding to at least a sequence portion of a tagging adapter of the at least one tagging adapter. The kit further comprises a reverse amplification primer comprising, from a 5' end to a 3' end, a second common sequence and a second sequence corresponding to at least a sequence portion of a tagging adapter of the at least one tagging adapter. DNA molecules of the tagged chromatin fragments can be amplified by means of the forward amplification primer and the reverse amplification primer to form a library of amplified DNA molecules.

The method and kit of the present invention can prepare a library of DNA molecules from a chromatin sample in a simple, cost-effective way that can be automated and takes less than one day to complete. The method and kit can be used for limited amounts of starting material, such as only one or a few hundred cells.

BRIEF DESCRIPTION OF THE DRAWINGS

The embodiments, together with further objects and advantages thereof, may best be understood by making reference to the following description taken together with the accompanying drawings, in which:

Fig. 1 illustrates an overview of the Restriction endonuclease Assisted Tagmentation Chromatin Immunoprecipitation (RAT-ChIP) method according to an embodiment.

Figs. 2A-2D - Restriction enzymes can be used for chromatin fragmentation. (2A) Restriction of chromatin using ten frequently cutting restriction endonucleases. (2B) Restriction of chromatin using CviKI-1 enzyme. (2C) Combining more restriction endonucleases in a single reaction results in more efficient chromatin fragmentation. (2D) 5 minute restriction time is sufficient to fragment majority of the chromatin.

Figs. 3A-3B - *In silico* analysis of four restriction endonuclease hg19 genomic recognition sites. (3A) Predicted fragment size distribution following restriction with AluI, SaqAI, MvaI and HinfI restriction endonucleases. (3B) Size distribution of 299 fragments that remain larger than 1,000 bp after *in silico* restriction with AluI, SaqAI, MvaI and HinfI restriction endonucleases.

Figs. 4A-4B - RAT-ChIP can identify histone H3K4me3 modification enrichments from 100 cells. (4A) Average yield of RAT-ChIP libraries from 100 and 1,000 cells after 16 rounds of PCR. (4B) RAT-ChIP enrichments of H3K4me3 at the promoters of glyceraldehyde 3-phosphate dehydrogenase (GAPDH), vacuolar protein sorting protein 29 (VPS29) and zinc finger protein 7 (ZNF7) genes compared to negative control region using 100 and 1,000 cells. A representative experiment is shown.

Figs. 5A-5C - RAT-ChIP enabled genome wide histone modification profiling from 100 cells. (5A) Agarose gel electrophoresis of DNA after chromatin treatment with a combination of restriction enzymes (middle lane) and after tagmentation (left lane). (5B) University of California Santa Cruz (UCSC) genome browser custom histone H3K4me3 and H3K27me3 tracks of RAT-ChIP-seq with 100 and 1,000 K562 cells in comparison with ENCODE data. (5C) Clustered global correlation heatmap of different published histone H3K4me3 and H3K27me3 datasets and RAT-ChIP-seq in K562 cells.

Figs. 6A-6D - RAT-ChIP can identify differences in histone modifications between cell-lines. (6A) UCSC genome browser custom histone H3K4me3 and H3K27me3 tracks of RAT-ChIP-seq with 100 and 1,000 cells in K562 and H1299 cells. (6B) Clustered global correlation heatmap of histone H3K4me3 and H3K27me3 datasets of K562 and H1299 cells. (6C) Heatmap of histone H3K4me3 signal in K562 and H1299 cells in 4kb region centered around the transcriptional start sites (TSS) of 300 genes with either cell type specific or common signal. (6D) Enriched terms of GREAT GO analysis of top 500 peaks differentially enriched between K562 and H1299 cells.

Figs. 7A-7C - RAT-ChIP can identify histone H3K4me3 and H3K27me3 modification profiles from bovine blastocysts. (7A) Custom UCSC tracks of histone H3K4me3 and H3K27me3 profiles in GAPDH gene locus in inner cell mass (ICM) and trophectoderm (TE) of blastocyst stage embryos. (7B) 6-way Venn diagram to show overlaps of genes from six published datasets that are upregulated in bovine blastocyst stage ICM (left) or TE (right). Below the Venn diagram is a summary of number of genes that overlap with a shown number of experiments. (7C) Custom UCSC tracks of histone H3K4me3 and H3K27me3 profiles in Nanog gene locus in morula, and ICM and TE of blastocyst stage mouse embryos (Liu et al, 2016).

Figs. 8A-8C - Histone H3K4me3 and H3K27me3 modification profiles of ICM and TE of blastocyst stage bovine embryos. (8A) Average histone H3K4me3 (upper panels) and H3K27me3 (lower panels) profiles around TSS of genes that are upregulated in ICM (gray line) or TE (black line) in ICM (panels on the left) and TE (panels on the right). (8B) UCSC genome browser custom histone H3K4me3 and H3K27me3 tracks of bovine blastocyst ICM and TE in NANOG gene region. (8C) UCSC genome browser custom histone H3K4me3 and H3K27me3 tracks of bovine blastocyst ICM and TE in CDX2 gene region.

Fig. 9 illustrates an overview of the RAT-ChIP method according to an embodiment used in combination with mRNA library preparation.

Fig. 10 is a flow chart illustrating a method of library preparation according to an embodiment.

Fig. 11 schematically illustrates tagging of a chromatin fragment and binding of amplification primers to the tagged chromatin fragment according to an embodiment.

DETAILED DESCRIPTION

The present embodiments generally relate to library preparation, and in particular to a method and a kit for library preparation from chromatin samples.

The present invention provides a significant improvement within the field of library preparation from chromatin samples that could be used, for instance, for sequencing. The method and kit of the present invention are capable of preparing a library from a chromatin sample comprising a very low amount of starting chromatin material, such as obtained from a low cell number sample. In fact, experimental data as presented herein indicates that the present invention can achieve an efficient library preparation with only or even less than 100 cells as starting material. This should be compared to many prior art ChIP-based methods for library preparation often requiring several hundred thousand or even millions of cells in the starting sample.

These improvements in being able to use low cell number starting material are, however, achievable even without the need for complex and laborious techniques or specific apparatus or equipment. The improvements are possible by reducing the number of experimental steps otherwise causing loss of material, including DNA purification and decrosslinking steps that are common among prior art ChIP methods.

The library prepared according to the present invention can be for various applications. The library is, however, particular suitable for next generation sequencing.

Fig. 10 is a flow chart illustrating a method of library preparation according to an embodiment. Reference is also made to Fig. 1 showing an overview of an implementation example of the method, denoted Restriction endonuclease Assisted Tagmentation Chromatin Immunoprecipitation (RAT-ChIP), and Fig. 11. The method comprises enzymatically digesting, in step S2, a chromatin sample 2 using at least one restriction enzyme 3 to form a plurality of chromatin fragments 4. At least a portion of the plurality of chromatin fragments 4 are immobilized in step S3 onto at least one solid support 5 comprising a respective affinity molecule 6 having affinity for a polypeptide 7 bound to DNA molecules 8 in the at least a portion of the plurality of chromatin fragments 4. A next step S4 comprises tagging the at least a portion of the plurality of chromatin fragments 4 immobilized onto the at least one solid support 5 in a tagmentation process using a transposase 9 and at least one tagging adapter 10, 11 to form tagged chromatin fragments 12 immobilized onto the at least one solid support 5. DNA molecules 8 of the tagged chromatin fragments 12 are amplified in step S5 in presence of a forward amplification primer 20 and a reverse amplification primer 30 to form a library of amplified DNA molecules 13. According to the embodiments, the forward amplification primer 20 comprises, from a 5' end 21 to a 3' end 22, a first common sequence 23 and a first sequence 25 corresponding to at least a sequence portion of a tagging adapter 10 of the at least one tagging adapter 10, 11. The reverse amplification primer

30 correspondingly comprises, from a 5' end 31 to a 3' end 32, a second common sequence 33 and a second sequence 35 corresponding to at least a sequence portion of a tagging adapter 11 of the at least one tagging adapter 10, 11.

The results of the method shown in Figs. 1 and 10 are thereby a library of DNA molecules 13 obtained from a chromatin sample 2. As a consequence, a library of sufficiently large number of such DNA molecules 13 can be obtained even from a minute chromatin sample 2, such as originating from a cell sample 1 containing only very few cells.

In the art, sonication has mainly been used to fragment a chromatin sample 2 into chromatin fragments 4. However, such a sonication step is generally not suitable for chromatin samples 2 from a low amount of cells as an efficient fragmentation of the chromatin sample 2 using sonication requires comparatively large sample volumes using standard equipment available in most laboratories. Such larger volumes in turn inevitably leads to dilution of the material, which is highly undesirable in cases with low number cell samples 1 and small chromatin samples 2 containing a comparatively low amount of chromatin molecules.

As a consequence, the fragmentation of the chromatin sample 2 is performed in step S2 using enzymatic digestions. In the art, usage of micrococcal nucleases (MNases) (EC 3.1.31.1) have been suggested for the purpose of chromatin fragmentation. MNase is an endo-exonuclease that preferentially digests single-stranded nucleic acids but is also active against double-stranded DNA and ribonucleic acid (RNA). However, such MNases are deemed not to be optimal since they digest the entire DNA sequence between nucleosomes in the chromatin sample, resulting in inefficient adapter insertion when these nucleosomes will be used for tagging in step S4. Accordingly, the enzymatic digestion in step S2 is, according to the embodiments, performed with at least one restriction enzyme 3. A restriction enzyme, also denoted restriction endonuclease or restrictase in the art, is an enzyme that cleaves DNA into fragments at or near specific recognition sites within the DNA molecule known as restriction sites.

Experimental data as presented herein indicates that restriction enzymes 3 can cut DNA only in between nucleosomes in a chromatin sample 2 to produce chromatin fragments 4.

According to the embodiments, the at least one restriction enzyme 3 recognize a recognition site of *N* bases in the DNA sequence of the chromatin sample 2 in between nucleosomes. In an embodiment, *N* is from 4 up to 8, such as 4, 5, 6, 7 or 8 bases. In a particular embodiment *N* = 4.

The present embodiments can use a restriction enzyme 3 that produces so-called sticky ends in the cut DNA molecule and/or a restriction enzyme 3 that produces so-called blunt ends in the cut DNA molecule.

In an embodiment, step S2 of Fig. 10 comprises enzymatically digesting the chromatin sample 2 using multiple, i.e., at least two, restriction enzymes 3 to form the plurality of chromatin fragments 4. Thus, in this embodiment at least two different restriction enzymes 3 are used to cut the DNA molecule of the chromatin sample 2. A reason for using multiple restriction enzymes 3 is to tailor the size of the DNA molecules 8 in the chromatin fragments 4. Thus, generally the more restriction enzymes 3 used in the enzymatic digestion of step S2, the shorter the DNA molecules 8 in the chromatin fragments 4. Hence, in some applications and depending on the particular restriction enzymes and in particular their respective recognition sites, the DNA molecules 8 in the chromatin fragments 4 may be too long, for instance for achieving an efficient sequencing of the DNA molecules 13 following amplification in step S5. This means that depending on the application of the library of amplified DNA molecules 13, and thereby the requirements of the length of the DNA molecules 13, one restriction enzyme or multiple restriction enzymes could be used in step S2.

In an embodiment, step S2 of Fig. 10 comprises enzymatically digesting the chromatin sample 2 using at least two restriction enzymes, preferably at least three restriction enzymes and more preferably at least four restriction enzymes. In another embodiment, step S2 of Fig. 10 comprises enzymatically digesting the chromatin sample 2 using one, two, three, four, five or six restriction enzymes.

The restriction enzyme(s) according to the invention may be selected from the group consisting of AluI, SaqAI, MvaI, HinfI, BsuRI, and CviKI-1. Preferably, the restriction enzyme(s) according to the invention are selected from the group consisting of AluI, SaqAI, MvaI, HinfI, and BsuRI. For instance, a mixture of AluI, SaqAI, MvaI and HinfI could be used in step S2.

The chromatin sample 2 could be any sample comprising chromatin. Non-limiting, but illustrative, examples of such chromatin samples include oocytes chromatin; chromatin from preimplantation embryos in different stages of development, for example chromatin from blastocyst stage embryos dissected into inner cell mass (ICM) and trophoectoderm (TE); chromatin from fluorescence-activated cell sorting (FACS) sorted cells, for example chromatin from hematopoietic stem cells or chromatin from cancer initiating cells.

In a preferred embodiment, the chromatin sample 2 is a small chromatin sample 2, such as a sample comprising the chromatin of no more than 1,000 cells, no more than 900 cells, no more than 800 cells, no more than 700 cell, no more than 600 cells, no more than 500 cells, no more than 400 cells, no more

than 300, no more than 200, no more than 100 cells, no more than 70 cells, or no more than 50 cells. Although the present invention has the significant advantage of usage with small chromatin samples 2, the present invention can, of course, also be used for library preparation from chromatin samples 2 comprising the chromatin of more than 1,000 cells.

In another embodiment, the chromatin sample 2 comprises the chromatin of or from between about 50 cells and about 10,000 cells, preferably between about 50 cells and about 1,000 cells, more preferably between about 70 cells and about 1,000 cells, still more preferably between about 70 cells and about 500 cells, even more preferably between about 100 cells and about 250 cells. Alternatively, the chromatin sample 2 comprises the chromatin of between about 50 cells and about 250 cells, preferably between about 70 cells and about 250 cells. In a particular embodiment, the chromatin sample 2 according to the invention comprises the chromatin of about 100 cells.

As used herein, the expression “about” refers to plus or minus 10 %, preferably to plus or minus 5 %. For example, about 100 cells refers to between 90 and 110 cells, preferably between 95 and 105 cells.

In an embodiment, the method comprises an additional step S1 as shown in Fig. 10. This step S1 then comprises lysing a cell sample 1 to form the chromatin sample 2. Hence, in this embodiment, the method starts with a sample 1 of cells that are lysed to release chromatin. This chromatin is then further processed in steps S2 to S5 to form the library of amplified DNA molecules 13.

The cell sample 1 could be any sample comprising cells, the chromatin of which is to be analyzed. Non-limiting, but illustrative, examples of such cell samples include oocytes; preimplantation embryos in different stages of development, for example blastocyst stage embryos dissected into ICM and TE; FACS sorted cells, for example hematopoietic stem cells or cancer initiating cells.

A significant advantage of the present invention over prior art ChIP methods is that the present embodiments can be used with small cell samples 1, even minute cell samples 1 comprising very few cells. Thus, in an embodiment, the cell sample 1 consists of no more than 1,000 cells, such as no more than 900 cells, no more than 800 cells, no more than 700 cell, no more than 600 cells, no more than 500 cells, no more than 400 cells, no more than 300, no more than 200, no more than 100 cells, no more than 70 cells, or no more than 50 cells.

Although the present invention has the significant advantage of usage with small cell samples 1, the present invention can, of course, also be used for library preparation from cell samples 1 comprising more than 1,000 cells.

In another embodiment, the cell sample 1 comprises between about 50 cells and about 10,000 cells, preferably between about 50 cells and about 1,000 cells, more preferably between about 70 cells and about 1,000 cells, still more preferably between about 70 cells and about 500 cells, even more preferably between about 100 cells and about 250 cells. Alternatively, the cell sample 1 comprises between about 50 cells and about 250 cells, preferably between about 70 cells and about 250 cells. In a particular embodiment, the chromatin sample 2 according to the invention comprises about 100 cells.

Steps S1 and S2 can be performed serially, i.e., first step S1 and then step S2 as schematically indicated in Fig. 10. In an alternative embodiment, steps S1 and S2 are performed in a single step. In such an embodiment, the cell sample 1 is preferably contacted with a lysis-restriction mixture comprising a nuclear lysis buffer and the at least one restriction enzyme 3 to form the plurality of chromatin fragments 4.

The nuclear lysis buffer is thereby capable of lysing the cell membranes and the nuclear membranes of the cells in the cell sample 1 to release the chromatin present therein so that the at least one restriction enzyme 3 present in the lysis-restriction mixture can access the DNA molecules of the chromatin sample 2 and enzymatically digest it into chromatin fragments 4.

Step S3 of Fig. 1 comprises immobilizing the chromatin fragments 4 onto at least one solid support 5 comprising a respective affinity molecule 6. The affinity molecule or molecules 6 has or have affinity for a respective polypeptide 7 bound to the DNA molecules 8 in the chromatin fragments 4. The at least one affinity molecule 6 can be any molecule having affinity for and thereby capability of binding to a polypeptide 7 in the chromatin fragments 3.

In an embodiment, the at least one affinity molecule 6 is at least one antibody 6 binding specifically to a polypeptide 7 bound to the DNA molecules 8 in the at least a portion of the plurality of chromatin fragments 4. In this embodiment, step S3 thereby comprises performing chromatin immunoprecipitation using at least one solid support 5 comprising a respective such antibody 6.

The specificity of an antibody, or indeed any affinity molecule, can be determined based on affinity and/or avidity. The affinity, represented by the equilibrium constant for the dissociation of an antigen with the

antibody (K_d), is a measure for the binding strength between an antigenic determinant and an antigen-binding site on the antibody. The lesser the value of K_d, the stronger the binding strength between the antigenic determinant and the antibody. Alternatively, the affinity can also be expressed as the affinity constant (K_a), which is $1/K_d$. As will be clear to the skilled person, affinity can be determined in a manner known per se, depending on the specific antigen of interest.

Avidity is the measure of the strength of binding between an antibody and the pertinent antigen. Avidity is related to both the affinity between an antigenic determinant and its antigen binding site on the antibody and the number of pertinent binding sites present on the antibody.

Typically, antibodies will bind to their antigen with a dissociation constant (K_d) of 10^{-5} to 10^{-12} moles/liter (M) or less, and preferably 10^{-7} to 10^{-12} M or less and more preferably 10^{-8} to 10^{-12} M, i.e. with an association constant (K_a) of 10^5 to 10^{12} M^{-1} or more, and preferably 10^7 to 10^{12} M^{-1} or more and more preferably 10^8 to 10^{12} M^{-1}.

Generally, any K_d value greater than 10^{-4} M (or any K_a value lower than 10^4 M^{-1}) is generally considered to indicate non-specific binding.

Preferably, an antibody or affinity molecule of the embodiments will bind to the polypeptide with an affinity less than 500 nM, preferably less than 200 nM, more preferably less than 10 nM, such as less than 5 nM.

Specific binding of an antibody to an antigen or antigenic determinant can be determined in any suitable manner known per se, including, for example, Scatchard analysis and/or competitive binding assays, such as radioimmunoassays (RIA), enzyme immunoassays (EIA) and sandwich competition assays, and the different variants thereof known per se in the art.

According to the embodiments, the antibody could be a monoclonal antibody or a polyclonal antibody.

In an embodiment, the affinity molecule is a fragment of an antibody. Non-limiting, but illustrative, examples of such antibody fragments can be selected from the group consisting of a single chain antibody, a Fv fragment, a scFv fragment, a Fab fragment, a F(ab')2 fragment, a Fab' fragment, a Fd fragment, a single-domain antibody (sdAb), a scFv-Fc fragment, a di-scFv fragment and a complementarity-determining region (CDR) region.

The affinity molecule of the embodiments is not limited to antibodies or fragments thereof. There are also other types of affinity molecules that can be used to bind to polypeptides in chromatin fragments. For instance, there are protein domains that specifically recognize certain histone posttranslational modifications. These could be used as affinity molecules. If the protein of interest is tagged one could use antibody or another affinity molecule, such as biotin-streptavidin to capture the protein of interest.

As mentioned above, the affinity molecule(s) has (have) affinity for polypeptide(s) bound to DNA molecules in the chromatin fragments. It is, however, also or alternatively possible to use affinity molecules having affinity for selected DNA sequences of the DNA molecules in the chromatin fragments. An example is methyl-CpG-binding domain (MBD) that has affinity for methylated DNA sequences.

The polypeptide 7, for which the affinity molecule 6, such as antibody, has affinity could be any polypeptide or protein 7 that is bound to DNA 8 in a chromatin sample 2. In an embodiment, the polypeptide 7 is a histone. A histone is a currently preferred example of polypeptide 7. However, the embodiments are not limited thereto. Also other chromatin bound polypeptides could be captured by the affinity molecule. Further examples of polypeptides include transcription factors, coactivators, corepressors, polymerases, chromatin remodellers and nucleases.

The at least one solid support 5, onto which the chromatin fragments 4 are immobilized in step S3 can be any solid support at which the affinity molecules 6, such as antibodies, can be bound or anchored. For instance, the solid support 5 could be a surface of a well of a microtiter plate; a membrane, such as a silica membrane or a carboxyl membrane; beads, such as silica beads, magnetic silica beads, carboxyl beads, or magnetic carboxyl beads; protein A or protein G beads, protein A or protein G magnetic beads; sepharose protein A or protein G beads; agarose protein A or protein G beads; streptavidin beads, etc.

In a particular embodiment, step S3 of Fig. 10 comprises immobilizing the at least a portion of the plurality of chromatin fragments 4 on beads 5, preferably magnetic beads, comprising the respective affinity molecule 6, preferably a respective antibody.

The solid support 5 may comprise a single type of affinity molecule 6 or may comprise multiple different affinity molecules 6 having affinity for different polypeptides 7 in the chromatin fragments 4. For instance, in the case of a well of a microtiter plate, each such well may comprise bound affinity molecules of a given type or multiple different affinity molecules may be bound to the surface of a single well. In the

case of beads as solid support all the beads may bind the same affinity molecule or the same mixture of affinity molecules. Alternatively, some of the beads may bind one type of affinity molecule or one mixture of affinity molecules whereas other beads may bind another type of affinity molecules or another mixture of affinity molecules.

For instance, Dynabeads with bound antibodies 6 could be used as solid support 5 according to the embodiments.

Step S4 comprises tagging the chromatin fragments 4 while immobilized onto the solid support 5 in a tagmentation process using a transposase 9 and at least one tagging adapter 10, 11.

Transposase (EC 2.7.7) is an enzyme that binds to the end of a transposon and catalyzes the movement of the transposon to another part of the genome by a cut and paste mechanism or a replicative transposition mechanism. Transposase Tn5 is a transposase having simultaneous tagging and fragmentation properties. Accordingly, in addition to tagging chromatin fragments 4, such a transposase 9 could further reduce the length of the DNA molecules 8 in the chromatin fragments 4 to achieve a length more suitable for the subsequent processing of the library of DNA molecules 13, such as sequencing of the DNA molecules 13. Other transposes than transposase Tn5 could be used including, for instance, Mu transposase and Tn7 transposase.

In an embodiment, a single type of tagging adapter 10 is tagged by the transposase 9 onto both ends of the DNA molecules 8 in the chromatin fragments 4. In other embodiments, different tagging adapters 10, 11 are used to thereby enable tagging one of the ends of the DNA molecules 8 with a first tagging adapter 10 and the other end of the DNA molecules 8 with a second tagging adapter 11. It is of course possible to have more than two types of tagging adapters 10, 11. For instance, at least one of the tagging adapters 10, 11 may comprise a unique molecular identifier (UMI) that is thereby unique for each tagging adapter molecule, which will be further described herein.

In an embodiment, step S4 of Fig. 10 comprises tagging the at least a portion of the plurality of chromatin fragments 4 immobilized onto the at least one solid support 5 in a tagmentation process using a transposase 9 and a first tagging adapter 10 and a second tagging adapter 11 to form tagged chromatin fragments 12 immobilized onto the at least one solid support 5 and comprising the first tagging adapter 10 and the second tagging adapter 11. In this embodiment, step S5 preferably comprises amplifying the DNA molecules 8 of the tagged chromatin fragments 12 in presence of the forward amplification primer

20 and the reverse amplification primer 30 to form the library of amplified DNA molecules 13. According to this embodiment, the forward amplification primer 20 comprises, from the 5' end 21 to the 3' end 22, the first common sequence 23 and the first sequence 25 corresponding to at least a 5-end sequence portion of the first tagging adapter 10. The reverse amplification primer 30 comprises, from the 5' end 31 to the 3' end 32, the second common sequence 33 and the second sequence 35 corresponding to at least a 5'-end sequence portion of the second tagging adapter 11.

In an embodiment, the first tagging adapter 10 has the following nucleotide sequence 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3' (SEQ ID NO: 1) and the second tagging adapter 20 has the following nucleotide sequence 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3' (SEQ ID NO: 2).

As briefly mentioned in the foregoing, at least one of the first tagging adapter 10 and the second tagging adapter 11 may comprise a UMI. In an embodiment, each first tagging adapter 10 comprises a respective specific UMI. In another embodiment, each second tagging adapter 11 comprises a respective specific UMI. In a further embodiment, each first tagging adapter 10 comprises a respective specific UMI and each second tagging adapter 11 comprises a respective specific UMI.

In an embodiment, the UMI is unique for a pair of the first tagging adapter 10 and the second tagging adapter 11 and each tagged chromatin fragment 12 comprises a respective UMI different from UMIs of other tagged chromatin fragments 12.

The UMI is a short sequence introduced into the DNA molecules 8 of the chromatin fragments 4 during the tagmentation process and is thereby present in the resulting library of amplified DNA molecules 13. The UMI serves to reduce the quantitative bias introduced by replication, i.e., the amplification in step S5.

In an embodiment, the UMI is a random $n_1n_2n_3...n_k$ sequence, wherein n_i, i=1...k, is one of A, T, C, G and U. In an embodiment, the parameter k is from 1 up to 12, preferably from 2 up to 8, such as 6. With k=6, 4,096 unique UMIs are possible using the nucleotides A, T, C and G. If more UMIs are required, the parameter k could, for instance, be equal to 8 giving 65,536 unique UMIs.

The amplification of the DNA molecules 8 of the tagged chromatin fragments 12 in step S5 is preferably performed as a polymerase chain reaction (PCR) amplification using a polymerase, such as a heat-stable DNA polymerase, for instance the Taq polymerase.

In an embodiment, at least one of the forward amplification primer 20 and the reverse amplification primer 30 comprises a barcode sequence 24, 34 as indicated in Fig. 11. In an embodiment, the forward amplification primer 20 comprises a barcode sequence 24. In another embodiment, the reverse amplification primer 20 comprises a barcode sequence 34. In a further embodiment, the forward amplification primer 20 comprises a barcode sequence 24 and the reverse amplification primer 20 comprises a barcode sequence 34.

In an embodiment, the barcode sequence 24, 34 is a *P* nucleotides sample-specific barcode sequence 24, 34. The parameter *P* is from 4 up to 16 nucleotides, preferably from 4 up to 10 nucleotides, and more preferably 8 nucleotides.

As indicated above, the barcode sequence 24, 34 is preferably sample specific, i.e., can thereby be used to identify the source, such as chromatin sample 2, cell sample 1 or patient, from which the amplified DNA molecules 13 originates.

In an embodiment, the first common sequence 23 of the forward amplification primer 20 is one of P5 sequence (5'-AATGATACGGCGACCACCGA-3', SEQ ID NO: 3) and a P7 sequence (5'-CAAGCAGAAGACGGCATACGAGAT-3', SEQ ID NO: 4). The second common sequence 33 of the reverse amplification primer 30 is the other of the P5 sequence and the P7 sequence.

In this case, the ILLUMINA® sequencing technology could be used to *in situ* sequence at least a portion of the amplified DNA molecules 13 by synthesis. In more detail, the amplified DNA molecules 13 are immobilized on a flow cell surface designed to present the amplified DNA molecules 13 in a manner that facilitates access to enzymes while ensuring high stability of surface bound DNA molecules 13 and low non-specific binding of fluorescently labeled nucleotides.

In an embodiment, the forward amplification primer 20 has the following general sequence layout 5'-P7-annealing_to_adapter_1-3' and the reverse amplification primer has the following general sequence layout 5'-P5-annealing_to_adapter_2-3'. In another embodiment, the forward amplification primer 20 has the following general sequence layout 5'-P5-annealing_to_adapter_1-3' and the reverse amplification

primer has the following general sequence layout 5'-P7-annealing_to_adapter_2-3'. In further embodiments, one but not both of the amplification primers 20, 30 has a barcode sequence, such as having a forward amplification primer 20 according to 5'-P7-annealing_to_adapter_1-3' or 5'-P5-annealing_to_adapter_1-3' and a reverse amplification primer 30 according to 5'-P5[i5]annealing_to_adapter_1-3' or 5'-P7[i7]annealing_to_adapter_1-3' or having a forward amplification primer 20 according to 5'-P7[i7]annealing_to_adapter_1-3' or 5'-P5[i5]annealing_to_adapter_1-3' and a reverse amplification primer 30 according to 5'-P5-annealing_to_adapter_1-3' or 5'-P7-annealing_to_adapter_1-3'. In other embodiments, both amplification primers 20, 30 have a respective barcode sequence. Thus, in an embodiment the forward amplification primer 20 has the following general sequence layout 5'-P7[i7]annealing_to_adapter_1-3' and the reverse amplification primer 30 has the following general sequence layout 5'-P5[i5]annealing_to_adapter_2-3'. In another embodiment, the forward amplification primer 20 has the following general sequence layout 5'-P5[i5]annealing_to_adapter_1-3' and the reverse amplification primer 30 has the following general sequence layout 5'-P7[i7]annealing_to_adapter_2-3'.

In these embodiments, P7 and P5 indicate the above identified P7 sequence and P5 sequence, respectively. [i7] and [i5] are in these embodiments barcode sequences 24, 34 and preferably correspond to the ILLUMINA® index 1 (i7) and index 2 (i5) sequences. In an embodiment, [i7] is selected from the group consisting of N701 (TAAGGCGA), N702 (CGTACTAG), N703 (AGGCAGAA), N704 (TCCTGAGC), N705 (GGACTCCT), N706 (TAGGCATG), N707 (CTCTCTAC), N708 (CAGAGAGG), N709 (GCTACGCT), N710 (CGAGGCTG), N711 (AAGAGGCA) and N712 (GTAGAGGA). In an embodiment, [i5] is selected from the group consisting of N501 (TAGATCGC), N502 (CTCTCTAT), N503 (TATCCTCT), N504 (AGAGTAGA), N505 (GTAAGGAG), N506 (ACTGCATA), N507 (AAGGAGTA) and N508 (CTAAGCCT).

Thus, in an embodiment, the forward amplification primer 10 has the following nucleotide sequence 5'-AATGATACGGCGACCACCGANNNNNNNNTCGTCGGCAGCGTCAGATGTG-3' (SEQ ID NO: 5), wherein NNNNNNNN is selected from the above mentioned group for [i7], i.e., one of N701 to N711, or is another type of barcode sequence. In another embodiment, the forward amplification primer 10 has the following nucleotide sequence 5'-CAAGCAGAAGACGGCATACGAGATNNNNNNNNTCGTCGGCAGCGTCAGATGTG-3' (SEQ ID NO: 6), wherein NNNNNNNN is selected from the above mentioned group for [i5], i.e., one of N501 to N508, or is another type of barcode sequence.

In an embodiment, the reverse amplification primer 20 has the following nucleotide sequence 5'-CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTCTCGTGGGCTCGGAGATGT-3' (SEQ ID NO: 7), wherein NNNNNNNN is selected from the above mentioned group for [i5], i.e., one of N501 to N508, or is another type of barcode sequence. In another embodiment, the reverse amplification primer 20 has the following nucleotide sequence 5'-AATGATACGGCGACCACCGANNNNNNNNGTCTCGTGGGCTCGGAGATGT-3' (SEQ ID NO: 8), wherein NNNNNNNN is selected from the above mentioned group for [i7], i.e., one of N701 to N711, or is another type of barcode sequence.

In an embodiment, the method does not comprise any DNA purification prior to or during steps S2-S5. Thus, the relaxed need for any DNA purification in the preparation of the library minimizes loss of DNA. This in turn allows the method to be used also for low cell number samples and low amount of chromatin. This means that the first DNA purification is preferably only performed after amplification in step S5, when loss of material is no longer an issue.

In an embodiment, the method does not comprise any sonication steps. In clear contrast, chromatin fragmentation is achieved according to the present invention in terms of enzymatic digestion using at least one restriction enzyme. As previously mentioned herein, chromatin fragmentation using sonication requires comparatively large sample volumes, which in turn would heavily dilute a chromatin sample originating from a low cell number sample. The considerable downscale of sample volumes obtainable according to the present invention has the additional advantage of requiring small amounts and concentrations of reagents, which reduce the cost for the method.

In an embodiment, the method does not comprise any proteinase K treatment step. Proteinase K is traditionally employed in many ChIP protocols to digest proteins in the sample following cell lysis. Proteinase K is, however, rather thermostable and, if present in the reaction mixture, could digest the polymerase used in the amplification step S5. This means that prior art protocols utilizing proteinase K has one or more additional steps to inactivate or remove proteinase K from the chromatin sample 2 or the chromatin fragments 4 prior to amplification of DNA molecules 8. The present embodiment thereby relaxes the need for such additional step(s).

In an embodiment, the method does not comprise any step of crosslinking the polypeptide 7 to the DNA molecules 8 in the chromatin fragments 4 and decrosslinking the crosslinked polypeptide 7 from the DNA

molecules 8. Such crosslinking and decrosslinking steps are common in prior art ChIP protocols but have the associated disadvantage of leading to DNA loss.

Steps S2 to S5, and preferably also step S1, i.e., steps S1 to S5, are preferably performed in a single reaction tube as schematically indicated in Fig. 1. This relaxes the need for transferring samples from different reaction tubes or vessel, where each such sample transfer is marred by loss of material.

The library of DNA molecules 13 obtainable according to the method of the present invention can be further processed in subsequent step or steps. For instance, the DNA molecules 13 can be sequenced in an optional step S6 as shown in Fig. 10. This step S6 thereby comprises sequencing at least a portion of the amplified DNA molecules 13 by addition of at least one sequencing primer having a sequence corresponding to or complementary to at least a portion of the at least one tagging adapter 10, 11.

In an embodiment, a single sequencing primer is used in step S6. However, if multiple tagging adapters 10, 11 are used in step S4 to tag the chromatin fragments 4, then multiple sequencing primers may be used in step S6. For instance, a first sequencing primer having a sequence corresponding to or complementary to at least a portion of the first tagging adapter 10 and a second sequencing primer having a sequence corresponding to or complementary to at least a portion of the second tagging adapter 11 can be used in step S6.

In a particular embodiment, the first sequencing primer mentioned above has the following nucleotide sequence 5'-GCCTCCCTCGCGCCATCAGAGATGTGTATAAGAGACAG-3'(SEQ ID NO: 9) and the second sequencing primer mentioned above has the following nucleotide sequence 5'-GCCTTGCCAGCCCGCTCAGAGATGTGTATAAGAGACAG-3' (SEQ ID NO: 10).

In a particular embodiment, step S6 comprises *in vitro* sequencing or *in situ* sequencing the at least a portion of the amplified DNA molecules 13 immobilized onto a solid support based on the first common sequence 23 and/or the second common sequence 33.

In such an embodiment, the solid support preferably comprises immobilized nucleotide sequences complementary to the first common sequence 23 and/or immobilized nucleotide sequences complementary to the second common sequence 33.

The *in situ* or *in vitro* sequencing of step S6 preferably comprises *in situ* or *in vitro* sequencing by synthesis of the at least a portion of the amplified DNA molecules 13.

For instance, if the first common sequence 23 is one the P5 and P7 sequence and the second common sequence 33 is the other of the P5 and P7 sequence, the ILLUMINA® sequencing technology could be used to *in situ* sequence at least a portion of the amplified DNA molecules 13 by synthesis.

Sequence By Synthesis (SBS) uses four fluorescently labeled nucleotides to sequence the amplified DNA molecules 13 on the flow cell surface in parallel. During each sequencing cycle, a single labeled deoxynucleoside triphosphate (dNTP) is added to the nucleic acid chain. The nucleotide label serves as a terminator for polymerization so after each dNTP incorporation, the fluorescent dye is imaged to identify the base and then enzymatically cleaved to allow incorporation of the next nucleotide. More information of the ILLUMINA® sequencing technology can be found in Technology Spotlight: ILLUMINA® Sequencing, the teaching of which is hereby incorporated by reference.

The library prepared according to the present embodiments can be useful for other purposes than sequencing. For instance, the DNA molecules of the library can be analyzed using quantitative PCR (qPCR).

Fig. 9 illustrates an overview of the RAT-ChIP method according to an embodiment used in combination with cDNA library preparation from messenger RNA (mRNA) molecules. Thus, in this approach not only a chromatin sample obtained following lysis of cells in a cell sample are processed as disclosed herein. In clear contrast, also mRNA molecules present in the cells and thereby released following cell lysis could be captured and amplified to prepare a library of complementary DNA (cDNA) molecules from the mRNA molecules. The processing of such mRNA molecules is outlined in the upper part of Fig. 9.

mRNA molecules comprise a polyA tail. This is utilized to immobilize the mRNA molecules using at least one solid support comprising a respective oligo d(T) molecule capable of hybridizing to the polyA tail. Thus, in an embodiment, the method comprises an additional step of immobilizing mRNA molecules present in the cell sample onto at least one solid support comprising a respective oligo d(T) molecule. The at least one solid could be selected from examples of solid support described above for the capture and immobilization of chromatin fragments.

The mRNA molecules immobilized onto the at least one solid support are then preferably reverse transcribed to form cDNA molecules. The reverse transcription is performed by the addition of a reverse transcriptase along with deoxynucleotide triphosphates. The reverse transcriptase synthesizes a complementary strand of DNA hybridized to the original mRNA strand immobilized to the at least one solid support.

In order to synthesize an additional DNA strand, the RNA of the hybrid strand may be digested using an enzyme such as RNase H or through alkali digestion method. After digestion of the RNA, a single stranded DNA (ssDNA) is left and because single stranded nucleic acids are hydrophobic, it tends to loop around itself. It is likely that the ssDNA forms a hairpin loop in the 3' end. From the hairpin loop, a DNA polymerase can then use it as a primer to transcribe a complementary sequence for the ss cDNA resulting in a dsDNA. The cDNA molecules can then be amplified using at least one amplification primer.

In another embodiment, mRNA molecules are captured onto at least one solid support using a respective oligo d(T) molecule that also comprises a universal or common sequence. Following cDNA synthesis, a primer with a common or universal 5'-end sequence and, for instance, GGG nucleotides in the 3' end can be used in template switching to add the universal or common sequence to the 3' end of the cDNA molecule. In a following, step the cDNA molecules can be amplified based on the universal or common sequences at the 5' and 3' ends to form a library of cDNA molecules, such as, for sequencing.

In a further embodiment, after first strand cDNA synthesis, a transposase can be added in a tagment the RNA:DNA heteroduplex immobilized onto the at least one solid support, see WO 2013/131962. The tagged nucleotide sequences can then be amplified as previously described herein for the RAT-ChIP method.

Another aspect of the embodiments relates to a kit for library preparation. This kit can thereby be used in the method of library preparation as disclosed herein. In an embodiment, the kit comprises at least one restriction enzyme 3 configured to enzymatically digest a chromatin sample 2 to form a plurality of chromatin fragments 4. The kit also comprises at least one solid support 5 comprising a respective affinity molecule 6 having affinity for a polypeptide 7 bound to DNA molecules 8 in the at least a portion of the plurality of chromatin fragments 4. The kit further comprises a transposase 9 and at least one tagging adapter 10, 11. The transposase 9 is configured to tag the at least a portion of the plurality of chromatin fragments 4 immobilized onto the at least one solid support 5 in a tagmentation process to form tagged chromatin fragments 12 immobilized onto the at least one solid support 5. The kit also comprises a

forward amplification primer 20 comprising, from a 5' end 21 to a 3' end 22, a first common sequence 23 and a first sequence 25 corresponding to at least a sequence portion of a tagging adapter 10 of the at least one tagging adapter 10, 11. The kit further comprises a reverse amplification primer 30 comprising, from a 5' end 31 to a 3' end 32, a second common sequence 33 and a second sequence 35 corresponding to at least a sequence portion of a tagging adapter 11 of the at least one tagging adapter 10, 11. DNA molecules 8 of the tagged chromatin fragments 12 can be amplified by means of the forward amplification primer 20 and the reverse amplification primer 30 to form a library of amplified DNA molecules 13.

In an embodiment, the kit comprises multiple restriction enzymes 3, such as at least two, preferably at least three and more preferably at least four of the restriction enzymes 3.

In an embodiment, the kit comprises restriction enzymes selected from the group consisting of AluI, SaqAI, MvaI, HinfI, BsuRI and CviKI-1.

In an embodiment, the kit comprises a nuclear lysis buffer configured to lyse a cell sample 1 to form the chromatin sample 2.

In an embodiment, the kit comprises a lysis-restriction mixture comprising the nuclear lysis buffer and the at least one restriction enzyme 3.

In an embodiment, the kit comprises at least one solid support 5 comprising a respective antibody 6 binding specifically to a polypeptide 7 bound to the DNA molecules 8 in the at least a portion of the plurality of chromatin fragments 4.

In an embodiment, the kit comprises beads 5, preferably magnetic beads, comprising the respective affinity molecule 6, preferably a respective antibody.

In an embodiment, the kit comprises a first tagging adapter 10 and a second tagging adapter 11. In this embodiment, the forward amplification primer 20 comprises, from the 5' end 21 to the 3' end 22, the first common sequence 23 and the first sequence 25 corresponding to at least a 5'-end sequence portion of the first tagging adapter 10 and the reverse amplification primer 30 comprises, from the 5' end 31 to the 3' end 32, the second common sequence 33 and the second sequence 35 corresponding to at least a 5'-end sequence portion of the second tagging adapter 11.

In an embodiment, at least one of the first tagging adapter 10 and the second tagging adapter 11 comprises a UMI.

In an embodiment, at least one of the forward amplification primer 20 and the reverse amplification primer 30 comprises a barcode sequence 24, 34.

In an embodiment, the first common sequence 23 is one of a P5 sequence (5'-AATGATACGGCGACCACCGA-3', SEQ ID NO: 3) and a P7 sequence (5'-CAAGCAGAAGACGGCATACGAGAT-3', SEQ ID NO: 4) and the second common sequence 33 is the other of the P5 sequence and the P7 sequence.

In an embodiment, the kit comprises at least one sequencing primer having a sequence corresponding to or complementary to at least a portion of the at least one tagging adapter 10, 11.

In an embodiment, the kit also comprises a reaction tube.

EXAMPLES

Chromatin immunoprecipitation coupled with next-generation sequencing (ChIP-seq) has revolutionized our understanding of chromatin related biological processes. The method, however, requires thousands of cells and has therefore limited applications in situations where cell numbers are limited such as highly pure cell populations or early developmental stages of mammalian preimplantation embryo where only few hundred cells might be available. Numerous attempts have been made to reduce the number of cells needed for successful ChIP-seq experiment, however, the developed methods are often complex, laborious or not sensitive enough.

Herein a new method denoted Restriction endonuclease Assisted Tagmentation Chromatin Immunoprecipitation (RAT-ChIP) is described. RAT-ChIP may obtain high quality genome-wide histone modification profiles from as few as 100 cells. The RAT-ChIP method is simple, cost-effective, takes one day to complete and can potentially be automated. The novel RAT-ChIP method was used to derive the first genome-wide maps of histone H3K4me3 and H3K27me3 modifications of inner cell mass (ICM) and trophoectoderm (TE) of bovine blastocyst stage embryos. Coupling the epigenetic landscape with published expression data showed that these modifications were relatively stable during the initial divergence of these two cell populations, however, there were examples such as the core pluripotency

factor NANOG, where transcriptional changes were accompanied by epigenetic alterations. The results showed that the novel RAT-ChIP method can be used to derive global histone post-translational modification profiles from a very low number of cells, making it especially suitable for studying early embryonic development.

Results

Development of RAT-ChIP method

Fig. 1 depicts the outline of an embodiment of the RAT-ChIP method. The whole protocol was designed to work essentially as a single tube – one day assay, reducing the number of necessary experimental steps, thus, minimizing the loss of material.

Initially when designing the assay, we wanted to take advantage of the simultaneous tagging and fragmentation properties of the Tn5 transposase. This would have achieved chromatin fragmentation and adapter insertion for sequencing library construction in one step. However, it appears that the transposase complex stays associated with both ends of the cut DNA (data not shown). We thus rethinked our strategy and decided to perform tagging of immunoprecipitated nucleosomal DNA on magnetic beads. Therefore, another means for chromatin fragmentation was needed. Sonication is the most commonly used method for this purpose but it was not a good option since it cannot be done in small volumes using the standard equipment available in most labs. Larger volumes in turn lead to dilution of the material, which is undesirable when working with small number of cells. Moreover, we wanted to avoid crosslinking, which due to harsh treatment needs to be done when using sonication, as it adds several additional steps to the protocol resulting in material loss.

Accordingly, enzymatic digestion methods, which can be performed in considerably smaller volumes without the need for crosslinking were used. MNase digestion that is used in native ChIP protocols for chromatin fragmentation was, however, not optimal, as MNase digests the entire free DNA between nucleosomes resulting in inefficient adapter insertion when these nucleosomes will be used for tagging. Alternative enzymatic means for chromatin digestion were therefore investigated. Restriction enzymes may be used for DNA footprinting to identify the locations of DNA bound proteins. Therefore, we tested whether restriction enzymes only could cut in between nucleosomes. Moreover, as their recognition sites are sequence specific, free DNA ends should remain available after digestion and could be used by transposase for adapter insertion. Theoretically even a single 4 bp recognizing restriction endonuclease should cut on average every 256 bp. However, the actual cutting frequency depends on the sequence and position of the nucleosomes. Nevertheless, a combination of frequent cutters should allow achieving

relatively even fragmentation coverage across the genome and fine enough resolution needed for histone modification profiling.

Accordingly, an array of frequently cutting restriction endonucleases were tested, see Table 1, for their ability to cut DNA in chromatin context. All used enzymes were from Thermo Scientific FastDigest lineup except for CviKI-1 from New England Biolavs, which allowed for rapid 5 minute digestions. Buffer conditions were used containing low amount of non-ionic detergents that could disrupt cell and nuclear membranes but did not interfere with the enzymatic activity. Out of the eleven tested restriction endonucleases six were able to cut chromatin with various efficiencies, evidenced by the appearance of clear nucleosomal ladders on agarose gel, see Figs. 2A and 2B.

Table 1 – restriction enzymes

Restriction enzyme	Recognition site
AluI	AGCT
BsuRI	GGCC
Csp6I	GTAC
FspBI	CTAG
HinfI	GANTC
HpyF3I	CTNAG
MspI	CCGG
MvaI	CCWGG
SaqAI	TTAA
TaqI	TCGA
CviKI-1	RGCY

AluI and SaqAI were the most efficient cutters followed by MvaI, HinfI, BsuRI and CviKI-1. Surprisingly, the other enzymes were not able to cut the chromatin even if they were able to cut naked DNA under the same conditions (data not shown). Although the best cutting single enzymes relatively efficiently fragmented chromatin, the majority of the DNA was still too big for ChIP. Chromatin was therefore cut using combinations of different restriction enzymes and a decrease in average DNA size was observed when more restriction endonucleases were used simultaneously, see Fig. 2C.

Next, the effect of incubation time on fragmentation efficiency was tested. There was no substantial variation in DNA size between 5, 10 and 15 min incubation times (Fig. 2D), indicating that 5 min was enough to digest majority of the DNA. The exception was CviKI-1, which is not a fast cutter and clearly fragments DNA better with increasing time (Fig. 2B).

Since restriction enzymes recognize specific DNA sequences *in silico* analysis was used to identify the genome-wide cutting sites of the used restriction endonucleases based on hg19 genome annotation. Using combination of four restriction endonucleases (AluI, SaqAI, MvaI and HinfI) 87 % of the genome was predicted to be cut into smaller pieces than 1,000 bp (Fig. 3A). In total there were 2,465 regions that based on the *in silico* analysis remained larger than 1,000 bp. Out of these 273 overlapped with gap regions, i.e., regions with no annotated sequence in hg19 genome version, 290 overlapped with so called black regions, i.e., regions that have abnormally high read counts in next-generation sequencing based studies, identified by ENCODE project and 2,065 overlapped with repeat regions downloaded from UCSC table browser RepeatMasker track. This left only 299 regions that did not overlap with any of the three lists (Fig. 3B), showing that at least *in silico* there are only a handful of unique genomic regions that cannot be effectively fragmented using a combination of restriction endonucleases.

Having identified that some restriction enzymes could be used for chromatin digestion, the RAT-ChIP protocol was tested. The use of enzymatic digestion enabled considerable downscale of the sample volumes, so that the digestions were performed in 1 µl and immunoprecipitation (IP) was performed in 11 µl final volume using 1 µl of magnetic beads pre-bound with antibody against histone modification of interest. After IP, beads were washed and tagging was performed directly on the beads. After another round of washes, the beads were used directly in polymerase chain reaction (PCR). To our knowledge, it is the first time that PCR is performed directly on the magnetic bead bound chromatin fragments. Skipping the decrosslinking, proteinase K treatment and DNA purification steps minimized the loss of DNA and allowed working with very low amounts of material. Even when starting with only 100 cells, enough material for sequencing was obtained after 16 rounds of PCR (Fig. 4A). Moreover, tagging on beads further decreased the fragment size, so that the whole RAT-ChIP library was between 200 and 500 bp in size (Fig. 5A), which is ideal for sequencing. Initial RAT-ChIP tests using 100 and 1,000 erythroleukemic K562 cells and H3K4me3 antibody followed by qPCR showed enrichment at the promoters of housekeeping genes glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and vacuolar protein sorting protein 29 (VPS29) compared to negative control regions in the end of zinc finger protein 7 (ZNF7) gene and an intergenic region in chromatin remodelling factor 17 (CHR17) (negative control)

(Fig. 4B) showing that using quantitative PCR (qPCR) the method was capable of detecting histone modification enrichments from only 100 cells.

RAT-ChIP enables high quality genome-wide histone modification profiling from 100 cells

The RAT-ChIP protocol was coupled with ILLUMINA® sequencing for genome-wide histone modification profiling. Human K562 cells derived from a chronic myelogenous leukemia for which many publicly available datasets exist was used. This allowed comparing the RAT-ChIP method with others. 100 and 1,000 cells were used with two different antibodies that recognize H3K4me3 and H3K27me3 to see if RAT-ChIP can be used to study both active and inactive histone marks. After paired-end sequencing the reads were mapped to hg19 genome, enrichment profiles were created and visualized as custom tracks in UCSC genome browser. Visual inspection and comparison to corresponding ENCODE data suggested that RAT-ChIP can produce high quality profiles that look similar to ENCODE data for both histone H3K4me3 and H3K7me3 modifications (Fig. 5B).

To further assess the quality of the RAT-ChIP data, it was compared to several other published ChIP-seq experiments in K562 cells. These included two datasets from ENCODE, a native ChIP-seq dataset (NCHIP), as well as two datasets from low input methods, see Table 2. Raw sequencing data was downloaded and all the datasets were processed in the same way. Comparing various parameters, such as % of mapped reads, GC% and fragment size for paired end data, showed that although there was variability between the compared datasets, RAT-ChIP performed comparably to other methods (Table 3). One parameter that, as expected, was clearly dependent on the number of cells used was the percentage of duplicated reads. In here, the low input methods clearly had more duplicated reads ranging from 46-64 %. In the K562 cell RAT-ChIP histone H3K4me3 dataset with 100 cells, 62 % of reads were duplicated. Initially, this may seem a lot but considering that it used 5 fold lower cell numbers compared to the Mint-ChIP dataset (which in turn is a pool from 5 different experiments to obtain more reads for comparison) and that there are less duplicates compared to ChIPmentation with 10,000 cells, RAT-ChIP performed on par with other low-input ChIP methods.

Table 2 - List of reanalyzed ChIP-seq datasets in K562 cells for comparison with RAT-ChIP-seq method

Name	GEO accession number	Histone modification	Type of experiment	No. of cells

ENCODE University of Washington rep1[a]	GSM945165	H3K4me3	Regular ChIP	millions
ENCODE University of Washington rep2[a]	GSM945165	H3K4me3	Regular ChIP	millions
ENCODE University of Washington rep1[a]	GSM945228	H3K27me3	Regular ChIP	millions
ENCODE University of Washington rep2[a]	GSM945228	H3K27me3	Regular ChIP	millions
ENCODE Bernstein lab rep1[b]	GSM733680	H3K4me3	Regular ChIP	millions
ENCODE Bernstein lab rep2[b]	GSM733680	H3K4me3	Regular ChIP	millions
ENCODE Bernstein lab rep1	GSM733658	H3K27me3	Regular ChIP	millions
ENCODE Bernstein lab rep2[b]	GSM733658	H3K27me3	Regular ChIP	millions
ChIPmentation[c]	GSM1782695	H3K4me3	ChIPmentation	1,0000
ChIPmentation[c]	GSM1782755	H3K4me3	ChIPmentation	500,000
ChIPmentation[c]	GSM1782693	H3K27me3	ChIPmentation	10,000
ChIPmentation[c]	GSM1782739	H3K27me3	ChIPmentation	500,000
Mint-ChIP[d]	GSM1918612-GSM1918616	H3K4me3	Mint-ChIP	5×500
Mint-ChIP[d]	GSM1918602-GSM1918606	H3K4me3	Mint-ChIP	5×1,000
Mint-ChIP[d]	GSM1918592-GSM1918596	H3K27me3	Mint-ChIP	5×500
Mint-ChIP[d]	GSM1918582-GSM1918586	H3K27me3	Mint-ChIP	5×1,000
nChIP[e]	GSM1141672	H3K4me3	native MNase ChIP	millions
nChIP[e]	GSM1141671	H3K27me3	native MNase ChIP	millions

a - Thurman et al, 2012; b – Dunham et al, 2012; c - Schmidl et al. 2015; d - van Galen et al, 2016; e - Mercer et al, 2013

Table 3 - Statistics and quality parameters of RAT-ChIP-seq and reanalyzed ChIP-seq datasets in K562 cells

Sample name	GEO	No. of cells	Modification	No. of reads
RAT-ChIP 100 cells		100	H3K4me3	31,456,882
RAT-ChIP 1,000 cells		1000	H3K4me3	69,563,299
Encode UW rep1	GSM945165	millions	H3K4me3	22,893,962
Encode UW rep2	GSM945165	millions	H3K4me3	20,872,420
Mint-ChIP 500 cells	(GSM1918612-GSM1918616)	5×500	H3K4me3	12,604,471
Mint-ChIP 1,000 cells	(GSM1918637-GSM1918641)	5×1,000	H3K4me3	13,469,262
ChIPmentation 10,000 cells	GSM1782695	10,000	H3K4me3	50,218,979
ChIPmentation 500,000 cells	GSM1782755	500,000	H3K4me3	36,095,458
Encode Bernstein rep1	GSM733680	millions	H3K4me3	17,903,857
Encode Bernstein rep2	GSM733680	millions	H3K4me3	30,270,485
nChIP	GSM1141672	millions	H3K4me3	11,639,776
RAT-ChIP 100 cells		100	H3K27me3	40,644,867
RAT-ChIP 1,000 cells		1,000	H3K27me3	52,698,558
Encode UW rep1	GSM945228	millions	H3K27me3	20,351,083
Encode UW rep2	GSM945228	millions	H3K27me3	20,647,926
Mint-ChIP 500 cells	(GSM1918592-GSM1918596)	5×500	H3K27me3	28,061,238
Mint-ChIP 1,000 cells	(GSM1918582-GSM1918586)	5×1,000	H3K27me3	34,926,999

ChIPmentation 10,000 cells	GSM1782693	10,000	H3K27me3	33,868,637
ChIPmentation 500,000 cells	GSM1782739	500,000	H3K27me3	45,743,646
Encode Bernstein rep1	GSM733658	millions	H3K27me3	23,651,262
Encode Bernstein rep2	GSM733658	millions	H3K27me3	28,165,277
nChIP	GSM1141671	millions	H3K27me3	13,528,824
Sample name	**No. of mapped reads**	**% of mapping**	**No. of duplicated reads**	**Duplication %**
RAT-ChIP 100 cells	29,159,974	92.7	19,516,178	62.04
RAT-ChIP 1,000 cells	65,054,480	93.52	37,576,090	54.02
Encode UW rep1	22,053,106	96.33	3,427,776	14.97
Encode UW rep2	20,048,358	96.05	2,723,832	13.05
Mint-ChIP 500 cells	11,644,094	92.38	6,397,058	50.75
Mint-ChIP 1,000 cells	12,423,818	92.24	6,242,445	46.35
ChIPmentation 10,000 cells	49,562,137	98.69	32,184,933	64.09
ChIPmentation 500,000 cells	35,657,898	98.79	8,365,134	23.18
Encode Bernstein rep1	14,079,366	78.64	1,678,365	9.37
Encode Bernstein rep2	23,040,669	76.12	3,347,238	11.06
nChIP	11,048,128	94.92	557,162	4.79
RAT-ChIP 100 cells	37,838,953	93.1	17,464,514	42.97

RAT-ChIP 1,000 cells	48,430,936	91.9	18,385,658	34.89
Encode UW rep1	16,293,012	80.06	6,345,994	31.18
Encode UW rep2	14,643,279	70.92	4,002,584	19.38
Mint-ChIP 500 cells	26,009,368	92.69	10,438,437	37.20
Mint-ChIP 1,000 cells	32,123,187	91.97	10,561,907	30.24
ChIPmentation 10,000 cells	33,397,012	98.61	5,801,869	17.13
ChIPmentation 500,000 cells	45,206,575	98.83	3,135,629	6.85
Encode Bernstein rep1	18,620,309	78.73	1,497,850	6.33
Encode Bernstein rep2	20,153,048	71.55	4,772,770	16.95
nChIP	9,670,065	71.48	484,226	3.58

Sample name	**Duplication rate**	**Read length**	**P25/Median/P75 insert size**	**GC%**
RAT-ChIP 100 cells	63.85	75 bp	132 / 186 / 265	45.54
RAT-ChIP 1,000 cells	47.62	75 bp	108 / 150 / 221	47.47
Encode UW rep1	12.38	36 bp		52.73
Encode UW rep2	10.89	36 bp		52.61
Mint-ChIP 500 cells	34.79	36 bp	200 / 260 / 330	52.48
Mint-ChIP 1,000 cells	33.17	36 bp	197 / 256 / 326	52.65
ChIPmentation 10,000 cells	60.74	50 bp		45.22
ChIPmentation 500,000 cells	18.87	50 bp		44.87

Encode Bernstein rep1	10.63	50 bp		47.28
Encode Bernstein rep2	12.5	50 bp		43.38
nChIP	3.15	36 bp		43.17
RAT-ChIP 100 cells	33.53	50 bp	108 / 168 / 265	42.74
RAT-ChIP 1,000 cells	18.52	50 bp	98 / 150 / 235	43.85
Encode UW rep1	37.03	36 bp		48.1
Encode UW rep2	25.78	36 bp		
Mint-ChIP 500 cells	25.97	36 bp	205 / 273 / 344	47.21
Mint-ChIP 1,000 cells	23.48	36 bp	201 / 269 / 340	47.08
ChIPmentation 10,000 cells	12.89	50 bp		41.94
ChIPmentation 500,000 cells	3.13	50 bp		40.94
Encode Bernstein rep1	6.39	50 bp		42.16
Encode Bernstein rep2	19.68	50 bp		47.12
nChIP	2.74	36 bp		41.08

Bigwig files were created and visualization in UCSC genome browser showed similar enrichment profiles for histone H3K4me3 modification in K562 cells for all the studied datasets with varying levels of background. As expected, in contrast to well-defined histone H3K4me3 peaks, histone H3K27me3 modification showed wide enrichment domains that can be several hundred kb-s long. Interestingly, the enrichment of H3K27me3 seemed to be relatively modest in majority of the studied datasets, with best background to noise signal seen with low input methods, first with Mint-ChIP followed by RAT-ChIP. In conclusion, based on visual inspection of enrichment profiles in UCSC genome browser, RAT-ChIP performed similar or better than many of the methods with much more input material.

To further assess how RAT-ChIP performed compared to other methods clustering was conducted based on global correlations in 10 kb windows between all the datasets. Two main clusters formed according to the studied histone H3 modification (Fig. 5C). Overall, the following correlations were found between H3K4me3 (r2=0.71-0.99) and H3K27me3 datasets (r2=0.48-0.93). Within the modifications, unsurprisingly, datasets from the same lab showed higher correlations and usually clustered together. In conclusion, this analysis showed that data produced using RAT-ChIP correlated similarly with other published datasets.

Regions enriched for histone H3K4me3 were determined using SICER and overlapped the regions between different samples. When one of the ENCODE datasets (UW1) was used as a reference RAT-ChIP H3K4me3 peaks overlapped with 72-74 % of the reference peaks, which was in the same range with Mint-ChIP (68-73 %) and NChIP (71 %) but lower than with other methods with more cells (82-90 %) and a replicate from the same lab (92 %). This analysis showed that despite lower overlap, low input methods were still capable of identifying the large majority of the enriched regions. Moreover, the regions not identified, had on average much lower signal in the original ENCODE data, suggesting that RAT-ChIP missed regions with low H3K4me3 enrichment.

RAT-ChIP can identify differences in histone modification profiles between cell-lines

Having identified that RAT-ChIP-seq data from K562 cells were comparable to other published datasets, it was next studied whether RAT-ChIP was capable of identifying cell type specific differences in histone modification profiles. To this end RAT-ChIP-seq was performed with 100 and 1,000 cells in human non-small cell lung carcinoma cell line H1299, for which no published ChIP-seq data existed, using histone H3K4me3 and H3K27me3 antibodies. After alignment and filtering, bigwig files were created and visualized in UCSC genome browser. Visual inspection and comparison of the RAT-ChIP tracks from H1299 and K562 cells showed similar enrichment profiles at the transcriptional start sites (TSS) of genes for H3K4me3 and broad H3K27me3 domains. Inspecting the loci of known hematopoietic transcription factors, such as GFI1b (Fig. 6A), GATA1, LMO2, ETO2 and entire globin locus, revealed clear differences between the two cell lines. For example, in GFI1b locus, there was enrichment of H3K4me3 around the TSS in K562 cells but the modification was completely absent in H1299 cells (Fig. 6A). The opposite was seen with H3K27me3, where a region around GFI1b gene had clearly higher signal of H3K27me3 in H1299 cells compared to K562 cells (Fig. 6A). Similar examples could be found for H1299 cell-line - several genes involved in epithelial to mesenchymal transition (EMT), such as TWIST2 and SIX1, showed the histone modification profile of active genes only in H1299 cells.

Pairwise correlation and clustering analysis showed that samples clustered first according to the profiled histone modifications and within the modifications according to cell-lines (Fig. 6B). To gain a more global view of the differences, differential H3K4me3 peaks were identified between the two cell lines. Fig. 6C shows heatmap of the signal intensities around TSS of 300 genes, which were differentially modified in one of the cell lines in contrast with 300 random genes that were not differentially modified. This analysis showed that RAT-ChIP-seq could identify hundreds of genes having different histone profiles between K562 and H1299 cells. To see if the differentially modified regions were near functionally relevant genes, a GO enrichment analysis was performed using GREAT. Analysis of 500 top H3K4me3 peaks that were more enriched in one of the cell lines compared to the other revealed enrichment of hematopoiesis related terms for K562 cells and signaling related terms for H1299 cells in biological processes category, confirming that RAT-ChIP detected differences occurred near genes with cell-type specific functions (Fig. 6D).

RAT-ChIP enabled histone profiling of blastocyst stage bovine embryos

Recently, bovine has been used as a model to study the molecular events that take place during early embryogenesis of large mammals – chromosomal instability in particular. RAT-ChIP-seq was thereby put to test and profile histone H3K4me3 and H3K27me3 modifications in blastocyst stage bovine embryos. This far only mouse embryos have been used for genome-wide histone profiling at such an early stage of development. Using *in vitro* fertilized embryos, micromanipulator was used in combination with laser microdissection to separate blastocysts into inner cell mass (ICM) and trophoectoderm (TE) fractions. Pooled material of three embryos was subsequently used for RAT-ChIP-seq experiments. Day 7-8 embryos comprise about 125 cells, 80 in TE and 45 in ICM. We pooled the TE cells and the ICM cells from 3 embryos and used them for 2 ChIPs, we thus used about 120 cells per ChIP for the TE experiments ((80×3)/2) and about 67 cells per ChIP for the ICM experiments ((45×3)/2). After alignment to bosTau8 genome, bigwig tracks with enrichment profiles were created. As expected, histone H3K4me3 was enriched mostly at promoter regions while histone H3K27me3 had broad domains of enrichment as exemplified by looking at the locus around housekeeping gene GAPDH (Fig. 7A), showing that RAT-ChIP could be used to obtain genome-wide histone modification profiles from early developmental stage embryos. With as little as 67 cells, results were still of very good quality, underlying that even less cells could be used with this new method.

To gain more global view of how these two histone marks acted in regulation of gene expression, the histone data was intersected with published gene expression data from ICM and TE of bovine

blastocysts. Five studies were found where gene expression profiles of ICM and TE were compared (Ozawa et al, 2012; Brinkhof et al, 2015; Nagatomo et al, 2015; Zhao et al, 2016; Hosseini et al, 2015). Two of them used RNA–seq and two others Affymetrix microarrays. In addition, one of the studies had done comparison with *in vivo* and *in vitro* derived blastocysts, making it in total six datasets. The lists of differentially expressed genes between ICM and TE were obtained for all the studies and intersected them. Overall, the overlap was relatively modest – there were only 6 and 0 gene(s) that were consistently upregulated in ICM and TE, respectively, in all datasets. The same numbers for at least 5 overlapping datasets were 28 and 5 and for at least 3 overlapping datasets 210 and 221 for ICM and TE, respectively (Fig. 7B). This analysis showed that there was a lot of variability and that the changes between TE and ICM at the transcriptome level were not huge at this early stage. In order to link gene expression to histone modification profiles, the genes that were differentially expressed at least in 3 datasets were taken and the average histone H3K4me3 and H3K27me3 profiles were profiled around 10kb regions around TSS. The genes that were upregulated in ICM showed higher histone H3K4me3 signal around TSS in ICM and higher H3K27me3 signal in TE. In contrast, genes that were upregulated in TE showed higher histone H3K4me3 signal around TSS in TE and higher H3K27me3 signal in ICM confirming that gene expression changes were on average mirrored by expected changes in histone modifications (Fig. 8A). Although on average the changes in histone modifications were in the expected direction, the changes were relatively modest – on a single gene level SICER managed to identify higher H3K4me3 on the TSS of 77 genes out of 210 (37 %) ICM upregulated genes in ICM and on the TSS of 82 genes out of 221 (37 %) for TE upregulated genes in TE. However, the same numbers for TSS where H3K4me3 levels were upregulated at least 2x were 17 (8 %) and 12 (5.4 %) for ICM and TE upregulated genes respectively.

To identify genes, which are potentially polycomb regulated, average H3K27me3 levels for ICM and TE upregulated genes were calculated for regions spanning the whole gene plus 2 kb upstream and ±2 kb of TSS. In part due to low number of reads, 6 (gene+2 kb upstream) and 17 (±2kb) genes were identified where the changes in H3K27me 3 levels were at least 4 times higher in TE compared to ICM for ICM upregulated genes and 23 (gene+2 kb upstream) and 24 (±2 kb) genes where the changes in H3K27me 3 levels were at least 4 times higher in ICM compared to TE for TE upregulated genes.

We next looked at individual genes known to be important in either ICM or TE specification. The promoter region of NANOG, a well-known pluripotency gene in embryonic stem (ES) cells was in our combined list of ICM upregulated genes and had H3K4me3 peak in ICM but not in TE (Fig. 8B). This was different from recently published data from mouse where NANOG promoter region was enriched for H3K4me3 in both

ICM and TE (Fig. 7C) (Liu et al, 2016). Moreover, despite of low read density NANOG locus also had more H3K27me3 in the NANOG locus in TE cells compared to ICM, which was not the case in the published mouse data (Fig. 7C).

In contrast to NANOG, the master regulator of TE development, CDX2, was in the list of ICM upregulated genes in only one of the expression datasets and was enriched for H3K4me3 in both cell types. Interestingly, there was an enrichment of H3K27me3 upstream of CDX2 gene specifically in ICM. The same region came up as the first hit when BLAT alignment was performed using a sequence of the recently characterized mouse trophectoderm specific enhancer upstream of CDX2 gene, suggesting that CDX2 might be down regulated through polycomb mediated enhancer repression. These examples showed that RAT-ChIP-seq data can be used for hypotheses generation to identify molecular mechanisms important for early embryonic development.

Discussion

A novel low input ChIP method called RAT-ChIP has been developed and can be used to create genome-wide histone modification profiles from only 100 cells.

There are several important modifications to the standard protocol that enabled achievement of successful results with such a low number of cells.

First - the use of restriction enzymes for chromatin fragmentation enabled usage of small sample volumes, which is important when working with low amount of starting material. The volumes used in other published protocols, use much higher volumes. In addition, due to small volumes the reagent costs were reduced significantly. For example, in a typical ChIP experiment 30 µl of magnetic beads are used compared to 1 µl in RAT-ChIP. Similar to MNase, restriction enzymes only cut in between nucleosomes but in contrast to MNase they leave DNA overhangs that can be used for sequencing library generation by tagmentation using chromatin as a template. A drawback of using restriction enzymes is that the cutting is not random. However, combining several frequently cutting enzymes enabled optimization of the coverage and desired fragment size. As restriction endonucleases are sequence specific and only cut in between nucleosomes there was no problem of over digestion. Moreover, due to the sequence specificity it was possible to predict genome-wide cutting sites. In combination of the four restriction endonucleases used in this study based on *in silico* analysis most of the genome was fragmented to the size suitable for ChIP. The larger fragments that remained often overlapped with gaps, ENCODE project identified black regions (regions that have abnormally high number of reads in next generation

sequencing data) or repeats that are difficult to analyze. Moreover, the larger regions that can cause false positive signals could be identified *in silico* and in case of need removed from further analysis.

Second – minimization of steps where material could be lost, such as centrifugations and DNA extractions. All steps in the protocol were carried out in a single tube so that the first DNA purification occurred only after PCR, when loss of material is not anymore an issue.

Third - simple, one step library preparation. In addition being extremely simple and cost effective, due to the random nature of tagmentation it allowed to further decrease DNA size, so that majority of the fragments in the final library came from single nucleosomes. In contrast to previous methods, such as ChIPmentation, PCR was performed directly on magnetic beads using the bound chromatin as a template. Skipping DNA purification avoided loss of material. The combination of these steps resulted in one of the most sensitive ChIP assays reported to date. The method was also very fast taking less than a day to complete.

Using K562 cells, the histone H3K4me3 and H3K27me3 profiles created using RAT-ChIP compared well to other published datasets demonstrating that it could be used to profile chromatin marks associated with both active and inactive genes. By profiling histone H3K4me3 and H3K27me3 modifications in H1299 cells, RAT-ChIP could identify differences between cell lines.

In addition to the vast potential bovine has in farming and biomedicine, it also serves as a good model system to study the molecular events that take place during early embryogenesis as its development is more similar to human compared to other common model organisms, such as mouse. Majority of the epigenetics experiments, genome-wide studies in particular, have been performed in mouse. Using RAT-ChIP the first genome wide histone H3K4me3 and H3K27me3 modification profiles of ICM and TE of blastocyst stage bovine embryos was created. Considering that day 7-8 embryos consist on average about 125 cells (80 in TE and 45 in ICM) and 3 embryos were pooled for the experiments and used for 2 ChIPs, only 70-120 cells were used per one immunoprecipitation.

Combined analysis of the histone modification data with lists of ICM and TE upregulated genes from published papers showed that gene expression changes were on average reflected by expected changes in histone modifications. Yet, due to low levels of enrichment in some regions, not completely pure cell populations and probably cellular heterogeneity, the changes at epigenetic level were not huge. However, looking at the histone modification profiles of factors with known importance either in ICM or TE function

evidence could be seen of the involvement of epigenetics in gene regulation. For example, a well-known pluripotency factor NANOG TSS had H3K4me3 only in ICM but not in TE. This was different from mouse blastocysts were the whole Nanog gene is covered with H3K4me3 both in ICM and TE (Fig. 7C). It was shown recently that there is loss of H3K4me3 on NANOG upon human embryonal carcinoma NT2/D1 cell differentiation towards neural progenitors. This shows that there might be important interspecies differences that need to be considered when drawing conclusions about regulation of specific genes. Moreover, the results that NANOG locus had more H3K27me3 in TE was in agreement with previous data as it has been shown that upon human ES cell differentiation there is an increase in H3K27me3 levels in the NANOG locus. Although Ezh2 has been shown to be involved in downregulation of Nanog expression upon ES cell differentiation also in mice H3K27me3 did not seem to be enriched at the Naonog locus in published mouse TE data. This discrepancy might be explained by locus specificity, interspecies differences, timing or cell purity used in ChIP experiments.

RAT-ChIP has been tested with histone modifications. However, RAT-ChIP can also be used to profile other chromatin bound proteins, including transcription factors. As the interaction of transcription factors is in general more labile, crosslinking step is usually used in ChIP. However, there is a recent protocol called ORGANIC ChIP, which demonstrated that transcription factors can be also immunoprecipitated without the need for crosslinking (Kasinathan et al, 2014).

In summary, a novel, simple, yet sensitive RAT-ChIP method has been developed and can be used to study genome-wide modifications in chromatin bound proteins, such as histones, even in a limited number of cells.

Materials and Methods

Cell lines

Human K562 and H1299 cells were grown in Iscove's Modified Dulbecco's Medium (IMDM) and Dulbecco's Modified Eagle's Medium (DMEM) (both from Naxo, Estonia) respectively, supplemented with 10 % of fetal bovine serum (FBS) and penicillin/streptomycin (Naxo) in the presence of 5 % CO_2 at 37°C.

Oocyte collection and in vitro maturation

All chemicals used for *in vitro* embryo production were purchased from IVF Limited T/A IVF Bioscience. Slaughterhouse derived ovaries were transported to the laboratory in a 0.9 % sterile NaCl solution within 4 h after slaughter at approximately 32-37°C and washed twice in a 0.9 % NaCl solution. Using a vacuum pump (Minitüb GmbH), cumulus oocyte complexes (COC) from follicles with a diameter of 2-8 mm were

aspirated. Grade 1 COCs were washed and matured in groups of 50 in 500 µl of *in vitro* maturation medium in four-well plates (Nunc). Oocytes were incubated at 38.5°C with humidified 5 % CO_2 in air for 22-24 h.

In vitro fertilization

Frozen-thawed semen from a Holstein bull ZIARD (id EE 13993023) was used to fertilize the matured oocytes. Oocytes and sperm were co-incubated in groups of 50 in 500 µl of Brackett-Oliphant *in vitro* culture (BO-IVM) media in four-well plates (Nunc) at 38.5°C with humidified 5 % CO_2 in air for 22-24 h.

In vitro cultivation

Zygotes were individually cultured in 60 µl droplets BO-IVC media for 8 days at 38.5°C, 5 % CO_2, 5 % O_2 and 90 % N_2 with maximum humidity. Embryos having reached blastocyst stage by day 8 were collected and used for laser assisted microdissection.

Laser assisted microdissection to obtain ICM and TE fractions

Integra 3 micromanipulator (Research Instruments Limited) equipped with Saturn 5 Active™ laser system was used to manually separate bovine blastocysts into ICM and TE fractions (of note – while manual dissection achieves to get pure populations of TE, small fraction of TE cells remain associated with the separated ICM mass). Separated fractions from three blastocysts were pooled and used for subsequent RAT-ChIP-seq experiments.

RAT-ChIP-(seq)

1 µl ProtG Dynabeads (Thermo Fisher Scientific) were bound with 0.25 µg of corresponding antibody (H3K4me3 (07-473, Millipore), H3K27me3 (07-449,Millipore)) in 5 µl of complete immunoprecipitation (IP) buffer (20 mM tris(hydroxymethyl)aminomethane-HCl (Tris-HCl), pH 7.4, 2 mM ethylenediaminetetraacetic acid (EDTA), 150 mM NaCl, 0.1 % Triton X-100) at room temperature (RT, 20-25°C) for 2 h followed by two washes using IP buffer.

The density of cultured K562 or H1299 cells was determined using haemocytometer. Cells were spun down and resuspended in phosphate-buffered saline (PBS) at a density 100 or 1,000 cells per 0.5 µl. Subsequently 0.5 µl of lysis-restriction mix (1 µl FD (FastDigest) buffer combined with 3.75 µl of 2x nuclear lysis buffer (20 mM TrisHCl, pH 7.4, 20 mM NaCl, 6 mM MgCl, 0.2 % NP-40)) and 0.25 µl 4x restriction enzyme mix (AluI #FD0014, SaqAI#FD2174, HinfI#FD0804, MvaI#FD0554, all FastDigest enzymes from Thermo Scientific) was added to the cells and incubated 15 min on ice and thereafter 5

min at 37°C. Next, 1 µl of 0.2 % sodium deoxycholate (NaDOC), 0.2 % TritonX-100 with protease inhibitors was added to the samples and incubated 15 min on ice after which 8 µl of IP buffer (20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 2 mM EDTA and 1 %TritonX-100) and 1 µl of ProtG Dynabeads (Thermo Fisher Scientific) pre-bound with corresponding antibody was added to the samples. Chromatin immunoprecipitation was performed at 4°C for four hours with end-over end mixing. After IP, beads were washed twice with 100 µl of following buffers: low salt washing buffer (0.1 % sodium dodecyl sulfate (SDS), 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl (pH 8.1)), high salt washing buffer, IP buffer and 20 mM TrisHCl, pH 7.4.

Tagging was performed by resuspending the beads in 2.5 µl of transposase mix (prepared by mixing 5 µl of 2x buffer with 4 µl mQ and 1 µl Transposase) (Illumina NEXTERA® DNA Sample Prep Preparation kit) and incubation of 1 min at 37°C. Beads were washed once with 100 µl of low salt washing buffer, once with 20 mM TrisHCl pH 7.4 and resuspended in 5 µl of 20 mM TrisHCl, pH 7.4.

16 cycles of PCR was performed using bead bound DNA as a template by mixing 5 µl of beads with 2.5 µl of 5 µM forward (SEQ ID NO: 5) and reverse primers (SEQ ID NO: 7) and 10 µl of 2x NEBNext enzyme. The following PCR program was used: (72°C 5 min, 98°C 2 min, 98°C 10 sec, 63°C 10 sec, 72°C 1 min, repeat steps 3-5 15 times, hold at 4°C). PCR products were purified with Agencourt RNA XP magnetic beads (Coulter Beckman), eluted in 10 µl of Tris-HCl pH 7.4 followed by DNA quantification using Nanodrop. The resulting library was subjected to 50 or 75 bp paired end Illumina sequencing using either HiSeq2500 or NextSeq platforms. Alternatively, the library was analyzed with q-RT-PCR using Applied Biosystems 7900HT real-time qPCR machine, HOT FIREPol® EvaGreen® qPCR Mix Plus (Solis BioDyne) and following primers: GAPDH_F CCCGTCCTTGACTCCCTAG (SEQ ID NO: 11), GAPDH_R CTGGTTCAACTGGGCACG (SEQ ID NO: 12); VPS29_F TCGCTACTTCCTGTTCTGCA (SEQ ID NO: 13), VPS29_R GATAGGGGCACGGTCCTC (SEQ ID NO: 14); ZNF7_F TACTGTTTCCTCGCCAGCTC (SEQ ID NO: 15), ZNF7_R GAGGCAAAGGAGACAAAGCA (SEQ ID NO: 16); Neg_cntrl_F CAAATGTGGTCACTAAGGCAAC (SEQ ID NO: 17), Neg_cntrl_R GTGACTCTCCTGGACCAACA (SEQ ID NO: 18).

RAT-ChIP-seq with bovine blastocysts was performed as described above except that after dissection the cells were collected in 3 µl of embryo medium. Lysis/restriction buffer was prepared by combining 4.75 µl of 10x NL (100 mM TrisHCl, pH 7.4, 100 mM NaCl, 30 mM $MgCl_2$) buffer, 4.75 µl of 10x FastDigest buffer and 0.5 µl of mix of 4x restriction endonucleases. 0.75 µl of lysis/restriction buffer was added to 3 µl of cells and incubated 15 min on ice and 5 min at 37°C. Next, 1 µl of mix of 0.5 % NaDOC, 0.5 %

Triton X-100 with protease inhibitor cocktail was added to the samples and incubated 15 min on after which 15 µl of IP buffer (20 mM Tris-HCl, pH 7.4, 2 mM EDTA, 150 mM NaCl, 0.1 % Triton X-100) was added and sample was divided into 2 tubes, 10 µl each for subsequent IP with Dynabeads bound with corresponding antibodies.

In silico analysis of restriction enzyme cutting sites

Human genome hg19 version GRCh37.p13 from EnsEMBL website (http://grch37.ensembl.org/) was used for the *in silico* analysis. The consensus sequence for each restriction endonuclease (AluI, SaqAI, HinfI, MvaI) was mapped onto each chromosome sequence and coordinates of the matches were recorded accordingly using custom written Perl scripts. Gap and repeat regions were downloaded from UCSC Table browser (Kent et al, 2002).

Used publicly available data

Raw data from following ChIP-seq datasets from GEO database were downloaded and reprocessed as described below: GSM945165, GSM945228, GSM733680, GSM733658, GSM1782695, GSM1782755, GSM1782693, GSM1782739, GSM1918612-GSM1918616, GSM1918602-GSM1918606, GSM1918592-GSM1918596, GSM1918582-GSM1918586, GSM1141671, GSM1141672. For Fig. 7C, bigwig files of following datasets - GSM2082690, GSM2082693, GSM2082696, GSM2082698, GSM2082701, GSM2082703 – were downloaded. For ICM and TE gene expression comparisons gene lists from the following publications were used (Ozawa et al, 2012; Brinkhof et al, 2015; Nagatomo et al, 2015; Zhao et al, 2016; Hosseini et al, 2015).

RAT-ChIP-seq and expression data analysis

Sequencing reads were mapped to hg19 or bosTau8 genome using Bowtie2 (Langmead & Salzberg, 2012) using following parameters -k 2 -N 1. Next, bam files were sorted and indexed using Samtools (Li et al, 2009) and bigwig files were created using deepTools (Ramírez et al, 2016). Blacklist regions were used for hg19 genome annotation created during ENCODE project to exclude them from further analysis (Dunham et al, 2012). Bigwig files were visualized in UCSC genome browser as custom tracks (Kent et al, 2002). Peak calling was done using SICER (Xu et al, 2014). Manipulation with genomic regions, correlation and clustering analysis and heatmap generation was done using various tools in the Cistrome Galaxy server (Liu et al, 2011; Blankenberg et al, 2010). Locus specific genomic enrichments of H3K27me3 were calculated using bwtools (Pohl & Beato, 2014). Venn diagrams were created using InteractiVenn (Heberle et al, 2015). All sequencing data has been deposited to Gene Expression Omnibus (GEO) database under the accession number GSE103734.

The embodiments described above are to be understood as a few illustrative examples of the present invention. It will be understood by those skilled in the art that various modifications, combinations and changes may be made to the embodiments without departing from the scope of the present invention. In particular, different part solutions in the different embodiments can be combined in other configurations, where technically possible. The scope of the present invention is, however, defined by the appended claims.

REFERENCES

Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A & Taylor J (2010) Galaxy: a web-based genome analysis tool for experimentalists. Curr. Protoc. Mol. Biol. Chapter 19: Unit 19.10.1-21

Brinkhof B, van Tol HT, Groot Koerkamp MJ, Riemers FM, IJzer SG, Mashayekhi K, Haagsman HP & Roelen BA (2015) A mRNA landscape of bovine embryos after standard and MAPK-inhibited culture conditions: a comparative analysis. BMC Genomics 16: 277

Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, Khatun J, Lajoie BR, Landt SG, Lee B-K, Pauli F, Rosenbloom KR, Sabo P, Safi A, Sanyal A, Shoresh N, et al (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489: 57–74

van Galen P, Viny AD, Ram O, Ryan RJH, Cotton MJ, Donohue L, Sievers C, Drier Y, Liau BB, Gillespie SM, Carroll KM, Cross MB, Levine RL & Bernstein BE (2016) A Multiplexed System for Quantitative Comparisons of Chromatin Landscapes. Mol. Cell 61: 170–80

Heberle H, Meirelles GV, da Silva FR, Telles GP & Minghim R (2015) InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. BMC Bioinformatics 16: 169

Hosseini SM, Dufort I, Caballero J, Moulavi F, Ghanaei HR & Sirard MA (2015) Transcriptome profiling of bovine inner cell mass and trophectoderm derived from in vivo generated blastocysts. BMC Dev. Biol. 15: 49

Kasinathan S, Orsi GA, Zentner GE, Ahmad K & Henikoff S (2014) High-resolution mapping of transcription factor binding sites on native chromatin. Nat. Methods 11: 203–9

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM & Haussler D (2002) The human genome browser at UCSC. Genome Res. 12: 996–1006

Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat. Methods 9: 357–9

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R & 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079

Liu T, Ortiz JA, Taing L, Meyer CA, Lee B, Zhang Y, Shin H, Wong SS, Ma J, Lei Y, Pape UJ, Poidinger M, Chen Y, Yeung K, Brown M, Turpaz Y & Liu XS (2011) Cistrome: an integrative platform for transcriptional regulation studies. Genome Biol. 12: R83

Liu X, Wang C, Liu W, Li J, Li C, Kou X, Chen J, Zhao Y, Gao H, Wang H, Zhang Y, Gao Y & Gao S (2016) Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. Nature 537: 558–562

Mercer TR, Edwards SL, Clark MB, Neph SJ, Wang H, Stergachis AB, John S, Sandstrom R, Li G, Sandhu KS, Ruan Y, Nielsen LK, Mattick JS & Stamatoyannopoulos JA (2013) DNase I-hypersensitive exons colocalize with promoters and distal regulatory elements. Nat. Genet. 45: 852–9

Nagatomo H, Akizawa H, Sada A, Kishi Y, Yamanaka K, Takuma T, Sasaki K, Yamauchi N, Yanagawa Y, Nagano M, Kono T, Takahashi M & Kawahara M (2015) Comparing spatial expression dynamics of bovine blastocyst under three different procedures: in-vivo, in-vitro derived, and somatic cell nuclear transfer embryos. Jpn. J. Vet. Res. 63: 159–71

Ozawa M, Sakatani M, Yao J, Shanker S, Yu F, Yamashita R, Wakabayashi S, Nakai K, Dobbs KB, Sudano MJ, Farmerie WG & Hansen PJ (2012) Global gene expression of the inner cell mass and trophectoderm of the bovine blastocyst. BMC Dev. Biol. 12: 33

Pohl A & Beato M (2014) bwtool: a tool for bigWig files. Bioinformatics 30: 1618–1619

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F & Manke T (2016) deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 44: W160-5

Schmidl C, Rendeiro AF, Sheffield NC & Bock C (2015a) ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. Nat. Methods 12: 963–965

Technology Spotlight: ILLUMINA® Sequencing, Illumina Sequencing Technology, Highest data accuracy, simple workflow, and a broad range of applications. (https://studylib.net/doc/8828848/illumina-sequencing-technology, retrieved on October 6, 2017)

Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, Garg K, John S, Sandstrom R, Bates D, Boatman L, Canfield TK, Diegel M, Dunn D, Ebersol AK, Frum T, et al (2012) The accessible chromatin landscape of the human genome. Nature 489: 75–82

Xu S, Grullon S, Ge K & Peng W (2014) Spatial clustering for identification of ChIP-enriched regions (SICER) to map regions of histone methylation patterns in embryonic stem cells. Methods Mol. Biol. 1150: 97–111

Zhao X-M, Cui L-S, Hao H-S, Wang H-Y, Zhao S-J, Du W-H, Wang D, Liu Y & Zhu HB (2016) Transcriptome analyses of inner cell mass and trophectoderm cells isolated by magnetic-activated cell sorting from bovine blastocysts using single cell RNA-seq. Reprod. Domest. Anim. 51: 726–35

CLAIMS

1. A method of library preparation, said method comprising the steps:

b) enzymatically digesting (S2) a chromatin sample (2) using at least one restriction enzyme (3) to form a plurality of chromatin fragments (4);

c) immobilizing (S3) at least a portion of said plurality of chromatin fragments (4) onto at least one solid support (5) comprising a respective affinity molecule (6) having affinity for a polypeptide (7) bound to DNA molecules (8) in said at least a portion of said plurality of chromatin fragments (4);

d) tagging (S4) said at least a portion of said plurality of chromatin fragments (4) immobilized onto said at least one solid support (5) in a tagmentation process using a transposase (9) and at least one tagging adapter (10, 11) to form tagged chromatin fragments (12) immobilized onto said at least one solid support (5); and

e) amplifying (S5) DNA molecules (8) of said tagged chromatin fragments (12) in presence of a forward amplification primer (20) comprising, from a 5' end (21) to a 3' end (22), a first common sequence (23) and a first sequence (25) corresponding to at least a sequence portion of a tagging adapter (10) of said at least one tagging adapter (10, 11) and a reverse amplification primer (30) comprising, from a 5' end (31) to a 3' end (32), a second common sequence (33) and a second sequence (35) corresponding to at least a sequence portion of a tagging adapter (11) of said at least one tagging adapter (10, 11) to form a library of amplified DNA molecules (13).

2. The method according to claim 1, wherein step b) comprises enzymatically digesting (S2) said chromatin sample (2) using at least two, preferably at least three and more preferably at least four restriction enzymes (3) to form said plurality of chromatin fragments (4).

3. The method according to claim 1 or 2, wherein step b) comprises enzymatically digesting (S2) said chromatin sample (2) using restriction enzymes (3) selected from the group consisting of AluI, SaqAI, MvaI, HinfI, BsuRI and CviKI-1.

4. The method according to any of the claims 1 to 3, further comprising a) lysing (S1) a cell sample (1) to form said chromatin sample (2).

5. The method according to claim 4, wherein said cell sample (1) consists of no more than 1,000 cells, preferably no more than 100 cells.

6. The method according to claim 4 or 5, wherein steps a) and b) are performed in a single step by contacting said cell sample (1) with a lysis-restriction mixture comprising a nuclear lysis buffer and said at least one restriction enzyme (3) to form said plurality of chromatin fragments (4).

7. The method according to any of the claims 1 to 6, wherein step c) comprises performing (S3) chromatin immunoprecipitation using at least one solid support (5) comprising a respective antibody (6) binding specifically to a polypeptide (7) bound to said DNA molecules (8) in said at least a portion of said plurality of chromatin fragments (4).

8. The method according to any of the claims 1 to 7, wherein step c) comprises immobilizing (S3) said at least a portion of said plurality of chromatin fragments (4) on beads (5), preferably magnetic beads, comprising said respective affinity molecule (6), preferably a respective antibody.

9. The method according to any of the claims 1 to 8, wherein

step d) comprises tagging (S4) said at least a portion of said plurality of chromatin fragments (4) immobilized onto said at least one solid support (5) in a tagmentation process using a transposase (9) and a first tagging adapter (10) and a second tagging adapter (11) to form tagged chromatin fragments (12) immobilized onto said at least one solid support (5) and comprising said first tagging adapter (10) and said second tagging adapter (11); and

step e) comprises amplifying (S5) said DNA molecules (8) of said tagged chromatin fragments (12) in presence of said forward amplification primer (20) comprising, from said 5' end (21) to said 3' end (22), said first common sequence (23) and said first sequence (25) corresponding to at least a 5'-end sequence portion of said first tagging adapter (10) and said reverse amplification primer (30) comprising, from said 5' end (31) to said 3' end (32), said second common sequence (33) and said second sequence (35) corresponding to at least a 5'-end sequence portion of said second tagging adapter (11) to form said library of amplified DNA molecules (13).

10. The method according to claim 9, wherein at least one of said first tagging adapter (10) and said second tagging adapter (11) comprises a unique molecular identifier (UMI).

11. The method according to claim 10, wherein said UMI is unique for a pair of said first tagging adapter (10) and said second tagging adapter (11) and each tagged chromatin fragment (12) comprises a respective UMI different from UMIs of other tagged chromatin fragments (12).

12. The method according to claim 10 or 11, wherein said UMI is a random $n_1n_2n_3...n_k$ sequence, wherein n_i, i=1…k, is one of A, T, C, G and U, and k is from 4 up to 12, preferably from 4 up to 10, such as 6.

13. The method according to any of the claims 1 to 12, wherein at least one of said forward amplification primer (20) and said reverse amplification primer (30) comprises a barcode sequence (24, 34).

14. The method according to claim 13, wherein said barcode sequence (24, 34) is a *P* nucleotides sample-specific barcode sequence (24, 34), wherein *P* is from 4 up to 16 nucleotides, preferably from 4 up to 10 nucleotides, more preferably 8 nucleotides.

15. The method according to any of the claims 1 to 14, wherein
said first common sequence (23) is one of a P5 sequence 5'-AATGATACGGCGACCACCGA-3', SEQ ID NO: 3 and a P7 sequence 5'-CAAGCAGAAGACGGCATACGAGAT-3', SEQ ID NO: 4; and
said second common sequence (33) is the other of said P5 sequence and said P7 sequence.

16. The method according to any of the claims 1 to 15, further comprising f) sequencing (S6) at least a portion of said amplified DNA molecules (13) by addition of at least one sequencing primer having a sequence corresponding to or complementary to at least a portion of said at least one tagging adapter (10, 11).

17. The method according to claim 16, wherein step f) comprises *in situ* sequencing (S6) said at least a portion of said amplified DNA molecules (13) immobilized onto a solid support based on said first common sequence (23) and/or said second common sequence (33).

18. The method according to any of the claims 1 to 17, wherein said method does not comprise any DNA purification prior to or during steps b), c), d) or e).

19. The method according to any of the claims 1 to 18, wherein said method does not comprise any sonication step nor any proteinase K treatment step.

20. The method according to any of the claims 1 to 19, wherein said method does not comprise any step of crosslinking said polypeptide (7) to said DNA molecules (8) and decrosslinking said crosslinked polypeptide (7) from said DNA molecules (8).

21. The method according to any of the claims 1 to 20, wherein steps b) to e) are performed in a single reaction tube.

22. The method according to any of the claims 1 to 21, further comprising immobilizing mRNA molecules present in said cell sample onto at least one solid support comprising a respective oligo d(T) molecule.

23. The method according to claim 22, further comprising reverse transcribing said mRNA molecules immobilized onto said at least one solid support to form cDNA molecules.

24. A kit for library preparation, said kit comprises:

at least one restriction enzyme (3) configured to enzymatically digest a chromatin sample (2) to form a plurality of chromatin fragments (4);

at least one solid support (5) comprising a respective affinity molecule (6) having affinity for a polypeptide (7) bound to DNA molecules (8) in said at least a portion of said plurality of chromatin fragments (4);

a transposase (9) and at least one tagging adapter (10, 11), wherein said transposase (9) is configured to tag said at least a portion of said plurality of chromatin fragments (4) immobilized onto said at least one solid support (5) in a tagmentation process to form tagged chromatin fragments (12) immobilized onto said at least one solid support (5); and

a forward amplification primer (20) comprising, from a 5' end (21) to a 3' end (22), a first common sequence (23) and a first sequence (25) corresponding to at least a sequence portion of a tagging adapter (10) of said at least one tagging adapter (10, 11); and

a reverse amplification primer (30) comprising, from a 5' end (31) to a 3' end (32), a second common sequence (33) and a second sequence (35) corresponding to at least a sequence portion of a tagging adapter (11) of said at least one tagging adapter (10, 11), wherein DNA molecules (8) of said tagged chromatin fragments (12) can be amplified by means of said forward amplification primer (20) and said reverse amplification primer (30) to form a library of amplified DNA molecules (13).

25. The kit according to claim 24, wherein said kit comprises at least two restriction enzymes (3), preferably at least three restriction enzymes (3) and more preferably at least four restriction enzymes (3).

26. The kit according to claim 24 or 25, wherein said kit comprises restriction enzymes selected from the group consisting of AluI, SaqAI, MvaI, HinfI, BsuRI and CviKI-1.

27. The kit according to any of the claims 24 to 26, wherein said kit comprises a nuclear lysis buffer configured to lyse a cell sample (1) to form said chromatin sample (2).

28. The kit according to claim 27, wherein said kit comprises a lysis-restriction mixture comprising said nuclear lysis buffer and said at least one restriction enzyme (3).

29. The kit according to any of the claims 24 to 28, wherein said kit comprises at least one solid support (5) comprising a respective antibody (6) binding specifically to a polypeptide (7) bound to said DNA molecules (8) in said at least a portion of the plurality of chromatin fragments (4).

30. The kit according to any of the claims 24 to 29, wherein said comprises beads (5), preferably magnetic beads, comprising said respective affinity molecule (6), preferably a respective antibody.

31. The kit according to any of the claims 24 to 30, wherein said kit comprises a first tagging adapter (10) and a second tagging adapter (11), wherein said forward amplification primer (20) comprises, from said 5' end (21) to said 3' end (22), said first common sequence (23) and said first sequence (25) corresponding to at least a 5'-end sequence portion of said first tagging adapter (10) and said reverse amplification primer (30) comprises, from said 5' end (31) to said 3' end (32), said second common sequence (33) and said second sequence (35) corresponding to at least a 5'-end sequence portion of said second tagging adapter (11).

32. The kit according to claim 31, wherein at least one of said first tagging adapter (10) and said second tagging adapter (11) comprises a unique molecular identifier (UMI).

33. The kit according to any of the claims 24 to 32, wherein at least one of said forward amplification primer (20) and said reverse amplification primer (30) comprises a barcode sequence (24, 34).

34. The kit according to any of the claims 24 to 33, wherein

said first common sequence (23) is one of a P5 sequence 5'-AATGATACGGCGACCACCGA-3', SEQ ID NO: 3 and a P7 sequence 5'-CAAGCAGAAGACGGCATACGAGAT-3', SEQ ID NO: 4; and

said second common sequence (33) is the other of said P5 sequence and said P7 sequence.

35. The kit according to any of the claims 24 to 34, wherein said comprises at least one sequencing primer having a sequence corresponding to or complementary to at least a portion of said at least one tagging adapter (10, 11).

36. The kit according to any of the claims 24 to 35, wherein said kit comprises a reaction tube.

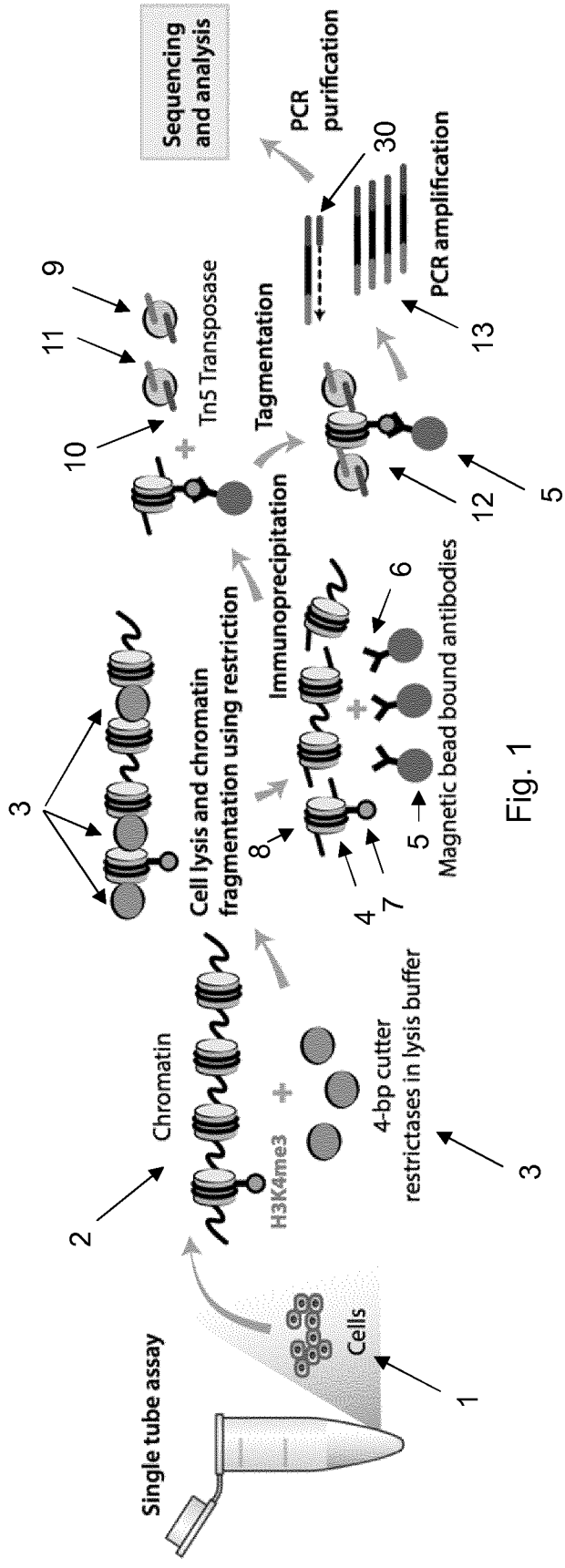

Single tube assay
Cells
1
Chromatin
H3K4me3
2
4-bp cutter
restrictases in lysis buffer
3
Cell lysis and chromatin
fragmentation using restriction
3
Immunoprecipitation
8
4
7
6
5
Magnetic bead bound antibodies
Tn5 Transposase
9
11
10
Tagmentation
12
5
PCR
purification
30
PCR amplification
13
Sequencing
and analysis

Fig. 1

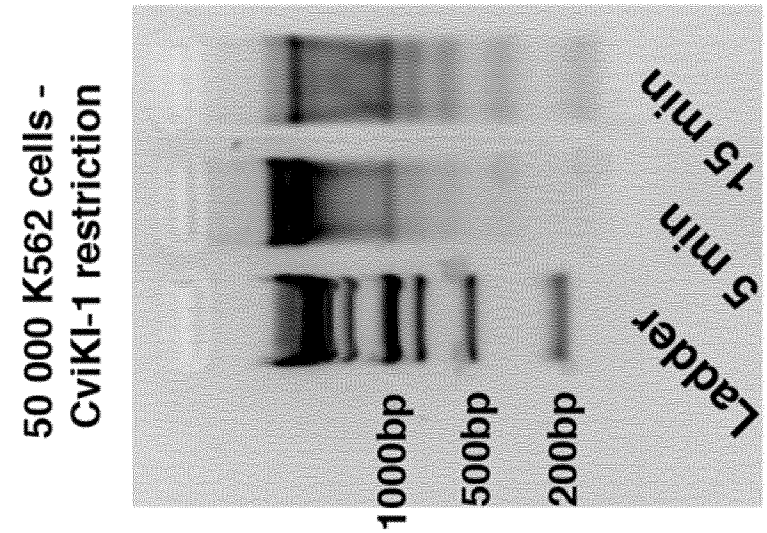

50 000 K562 cells - CviKI-1 restriction
1000bp
500bp
200bp
Ladder
5 min
15 min

Fig. 2B

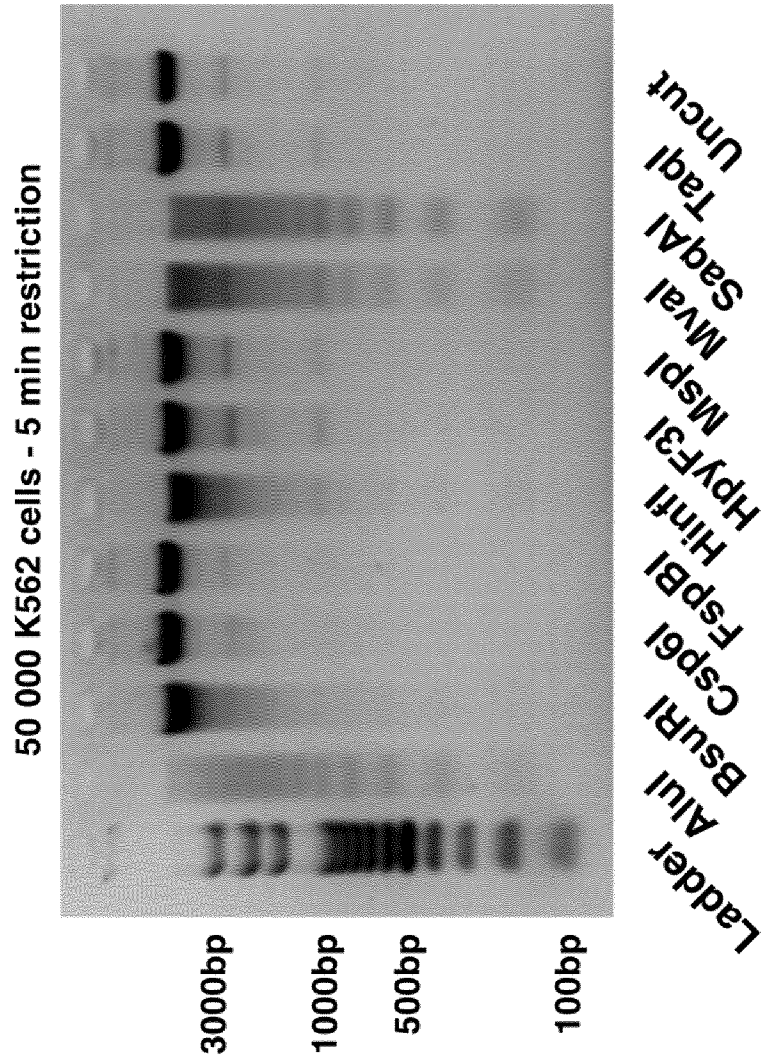

50 000 K562 cells - 5 min restriction
3000bp
1000bp
500bp
100bp
Ladder
AluI
BsuRI
Csp6I
FspBI
HinfI
HpyF3I
MspI
MvaI
SaqAI
TaqI
Uncut

Fig. 2A

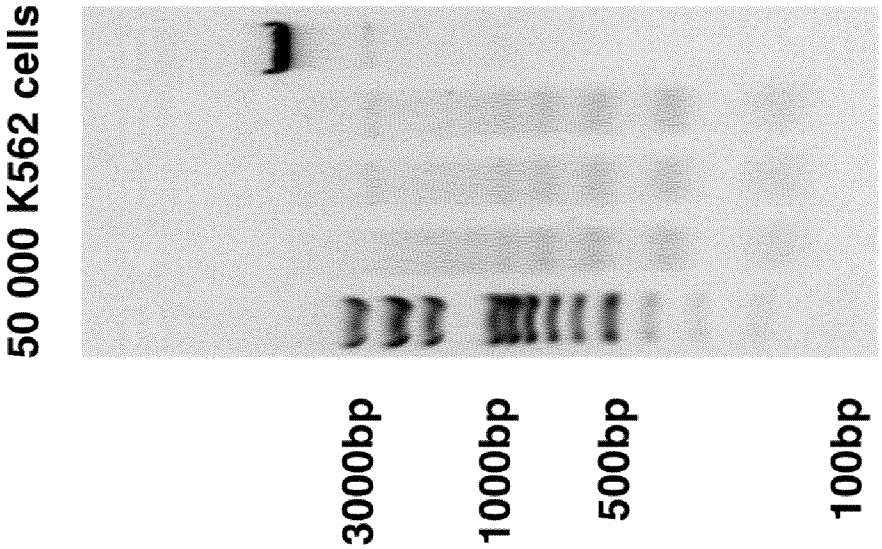

50 000 K562 cells
Ladder
5 min
10 min
15 min
Uncut
3000bp
1000bp
500bp
100bp

Fig. 2D

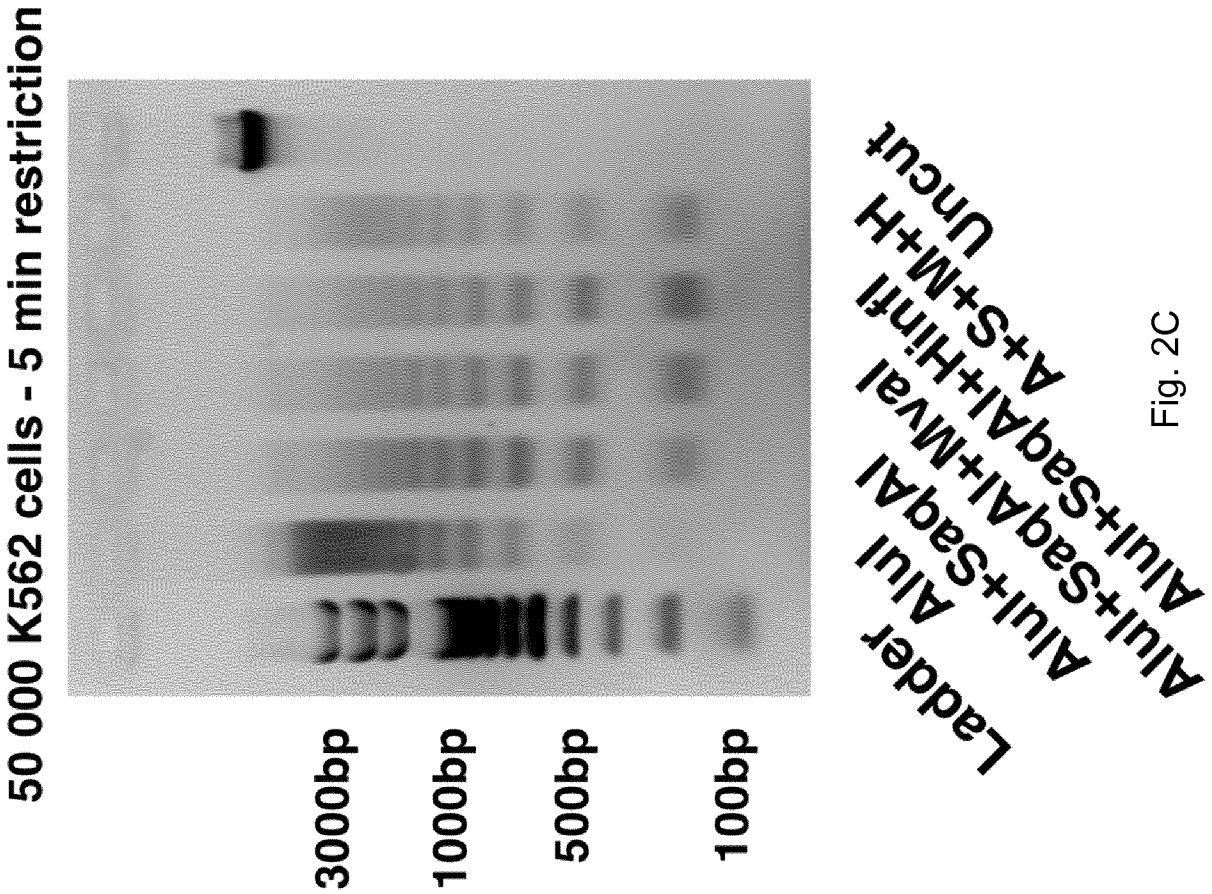

50 000 K562 cells - 5 min restriction
Ladder
AluI
AluI+SaqAI
AluI+SaqAI+MVaI
AluI+SaqAI+HinfI
AluI+SaqAI+M+H
A+S+M+H
Uncut
3000bp
1000bp
500bp
100bp

Fig. 2C

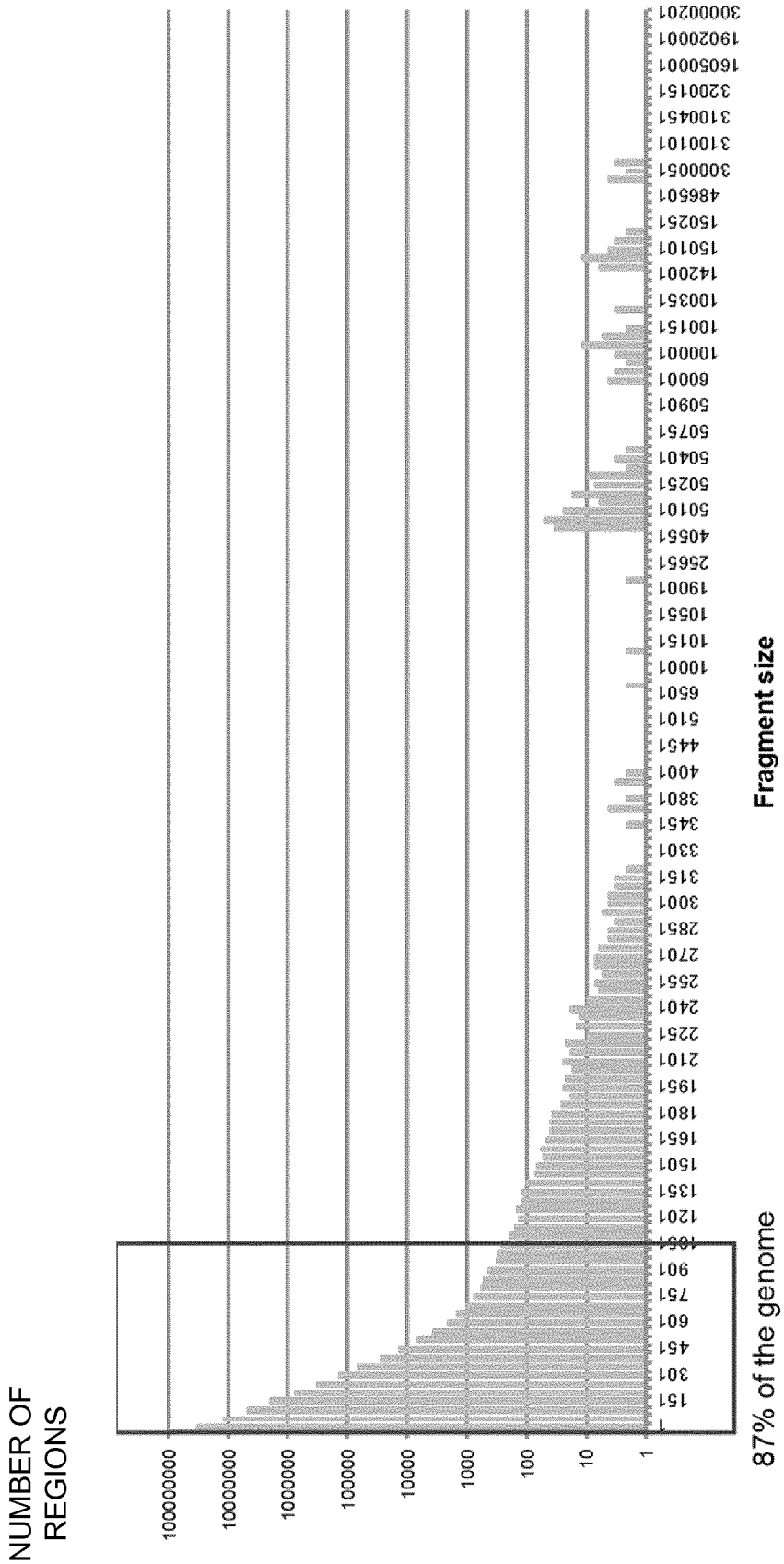

NUMBER OF REGIONS
100000000
10000000
1000000
100000
10000
1000
100
10
1
87% of the genome
Fragment size

Fig. 3A

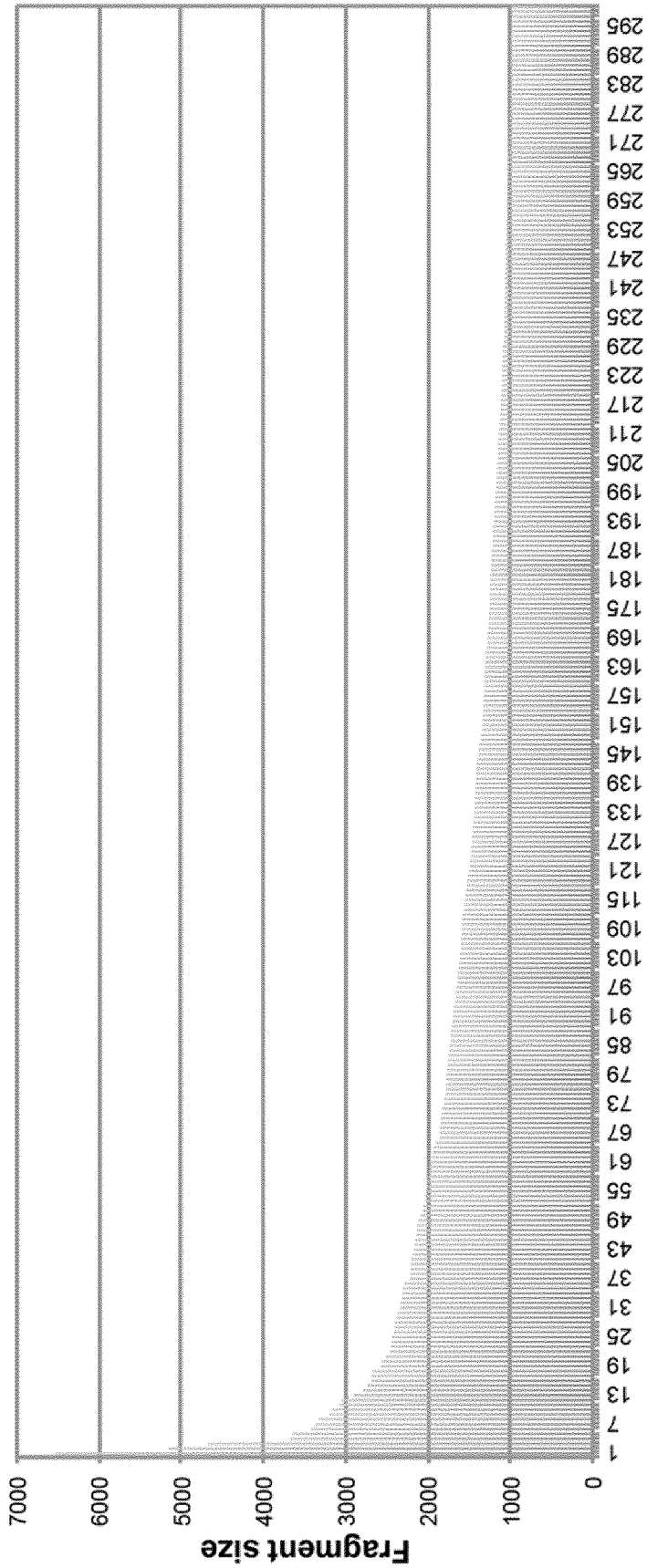

Fragment size
7000
6000
5000
4000
3000
2000
1000
0
Fragments
1
7
13
19
25
31
37
43
49
55
61
67
73
79
85
91
97
103
109
115
121
127
133
139
145
151
157
163
169
175
181
187
193
199
205
211
217
223
229
235
241
247
253
259
265
271
277
283
289
295

Fig. 3B

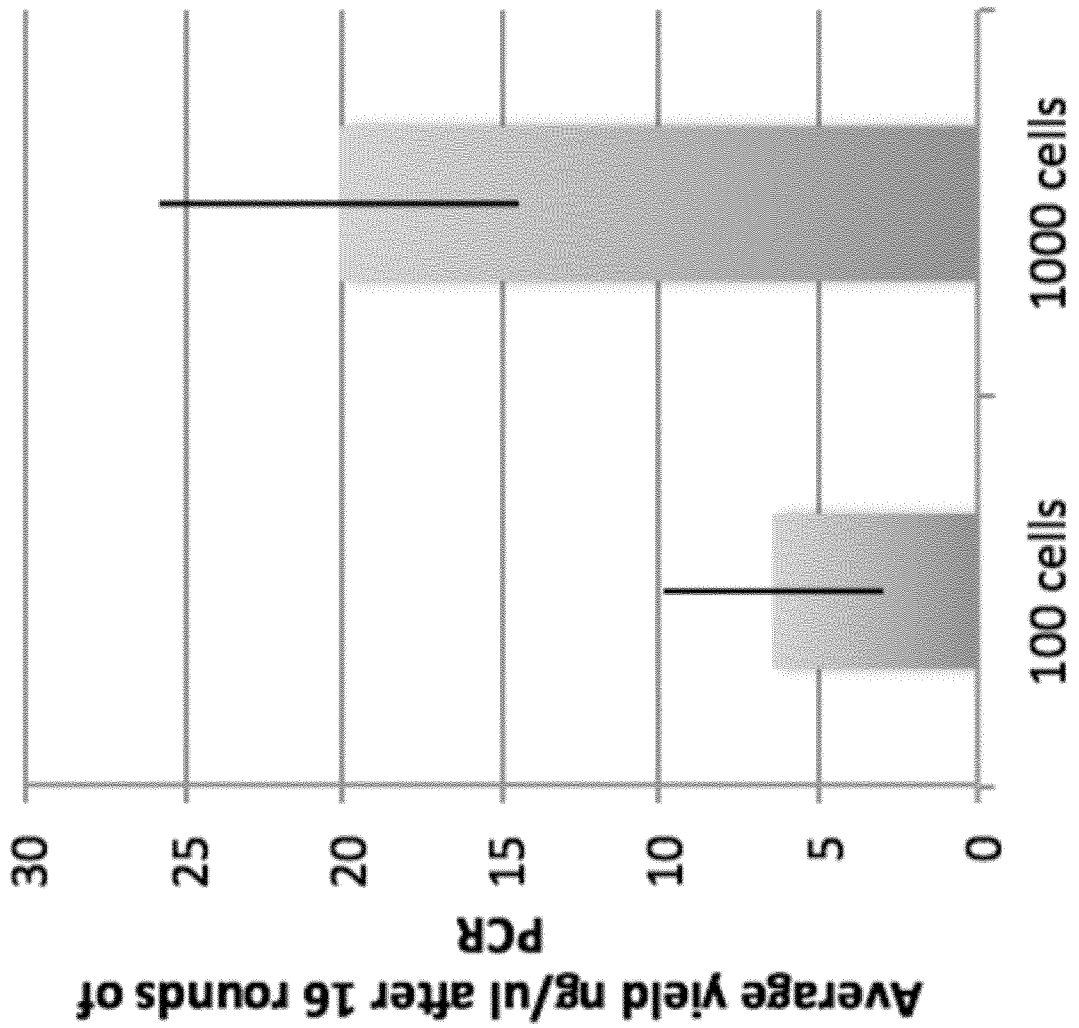

Average yield ng/ul after 16 rounds of PCR
30
25
20
15
10
5
0
100 cells
1000 cells

Fig. 4A

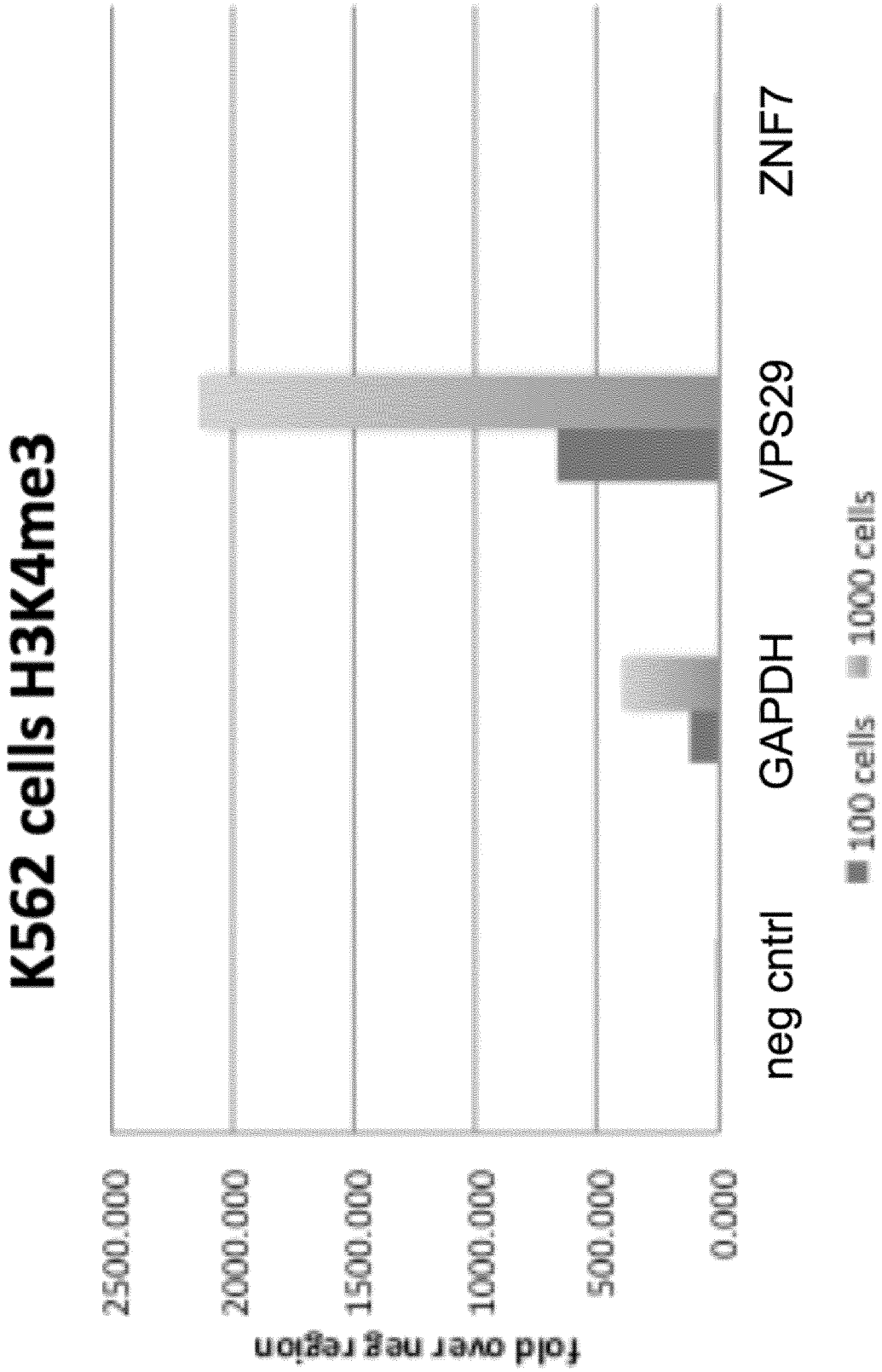

K562 cells H3K4me3
fold over neg region
2500.000
2000.000
1500.000
1000.000
500.000
0.000
neg cntrl
GAPDH
VPS29
ZNF7
100 cells
1000 cells

Fig. 4B

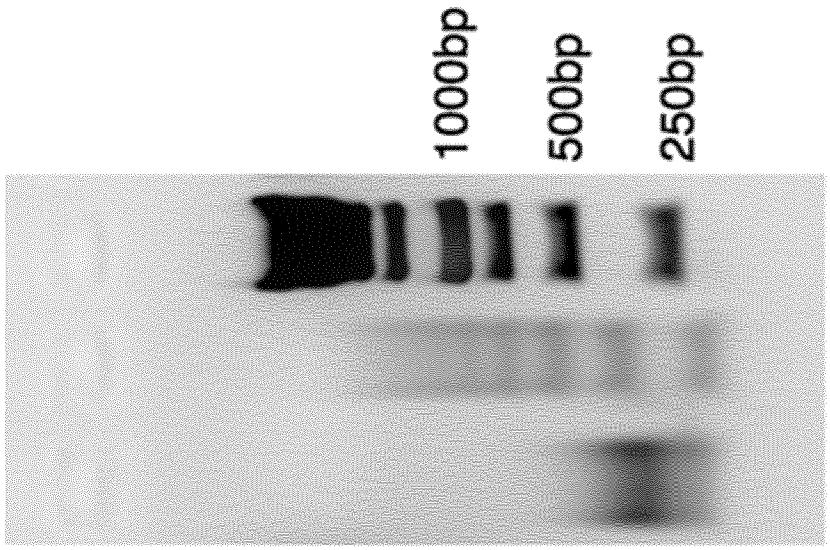


Fig. 5A

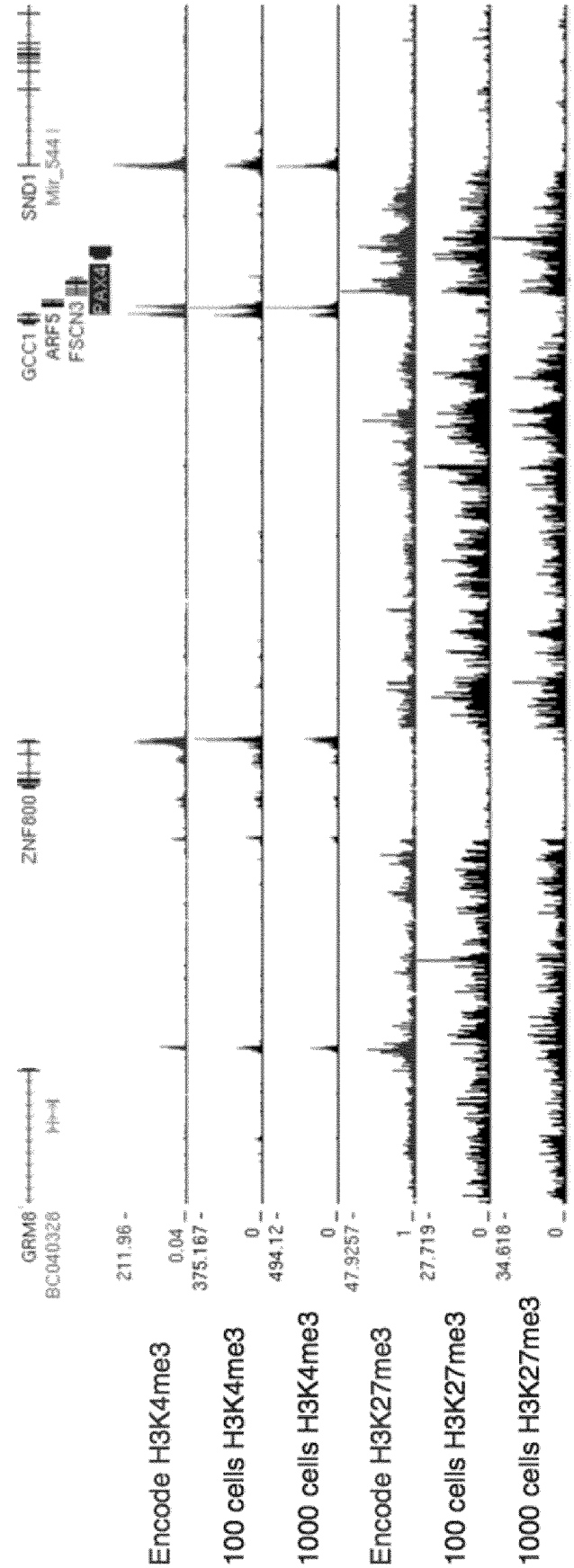

GRM8
BC040328
ZNF800
GCC1
ARFS
FSCN3
SND1
Mir_544
Encode H3K4me3
100 cells H3K4me3
1000 cells H3K4me3
Encode H3K27me3
100 cells H3K27me3
1000 cells H3K27me3
211.96
0.04
375.167
0
494.12
0
47.9257
1
27.719
0
34.618
0

Fig. 5B

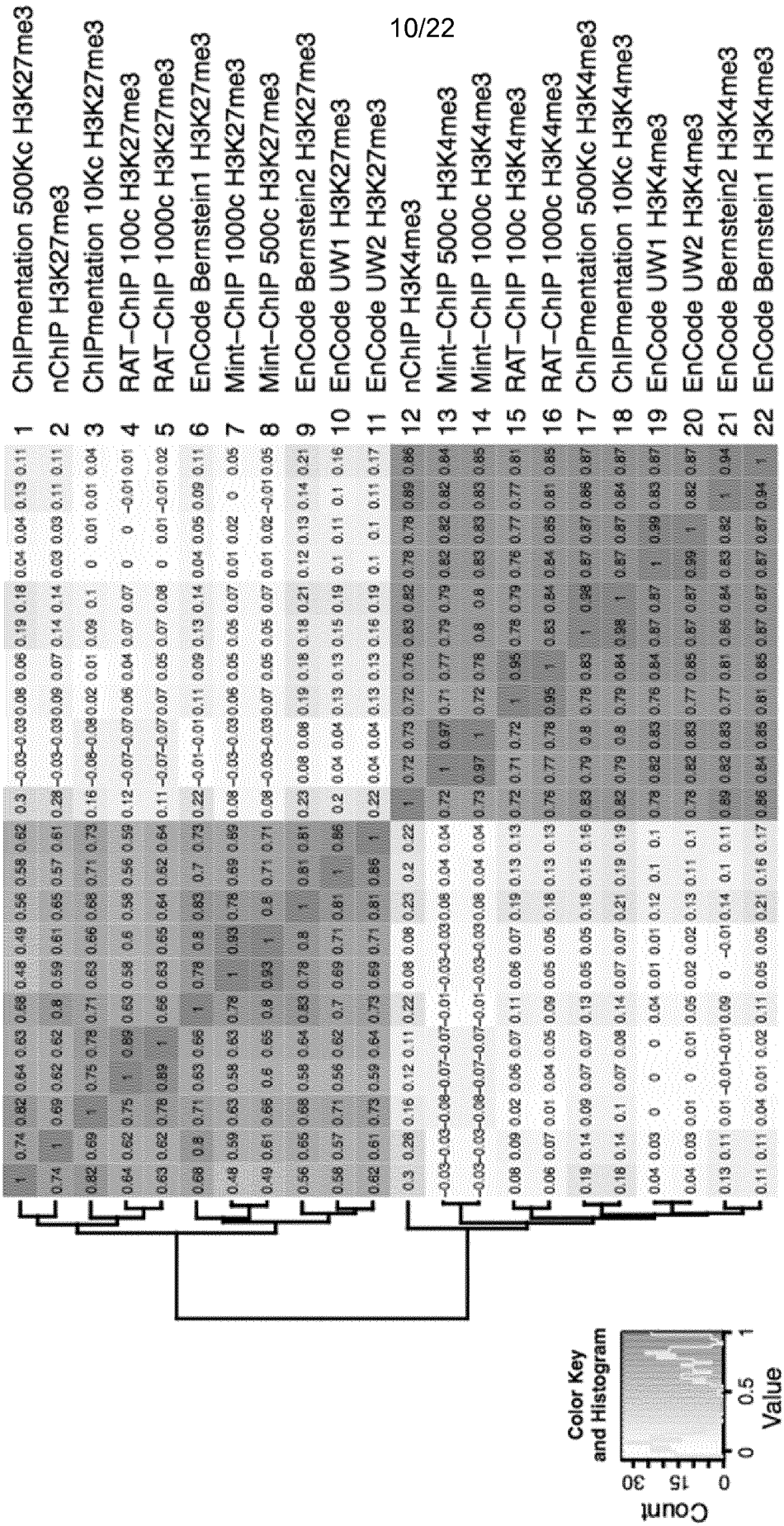

1 ChIPmentation 500Kc H3K27me3
2 nChIP H3K27me3
3 ChIPmentation 10Kc H3K27me3
4 RAT–ChIP 100c H3K27me3
5 RAT–ChIP 1000c H3K27me3
6 EnCode Bernstein1 H3K27me3
7 Mint–ChIP 1000c H3K27me3
8 Mint–ChIP 500c H3K27me3
9 EnCode Bernstein2 H3K27me3
10 EnCode UW1 H3K27me3
11 EnCode UW2 H3K27me3
12 nChIP H3K4me3
13 Mint–ChIP 500c H3K4me3
14 Mint–ChIP 1000c H3K4me3
15 RAT–ChIP 100c H3K4me3
16 RAT–ChIP 1000c H3K4me3
17 ChIPmentation 500Kc H3K4me3
18 ChIPmentation 10Kc H3K4me3
19 EnCode UW1 H3K4me3
20 EnCode UW2 H3K4me3
21 EnCode Bernstein2 H3K4me3
22 EnCode Bernstein1 H3K4me3
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Color Key and Histogram
Count
0 15 30
0 0.5 1
Value

Fig. 5C

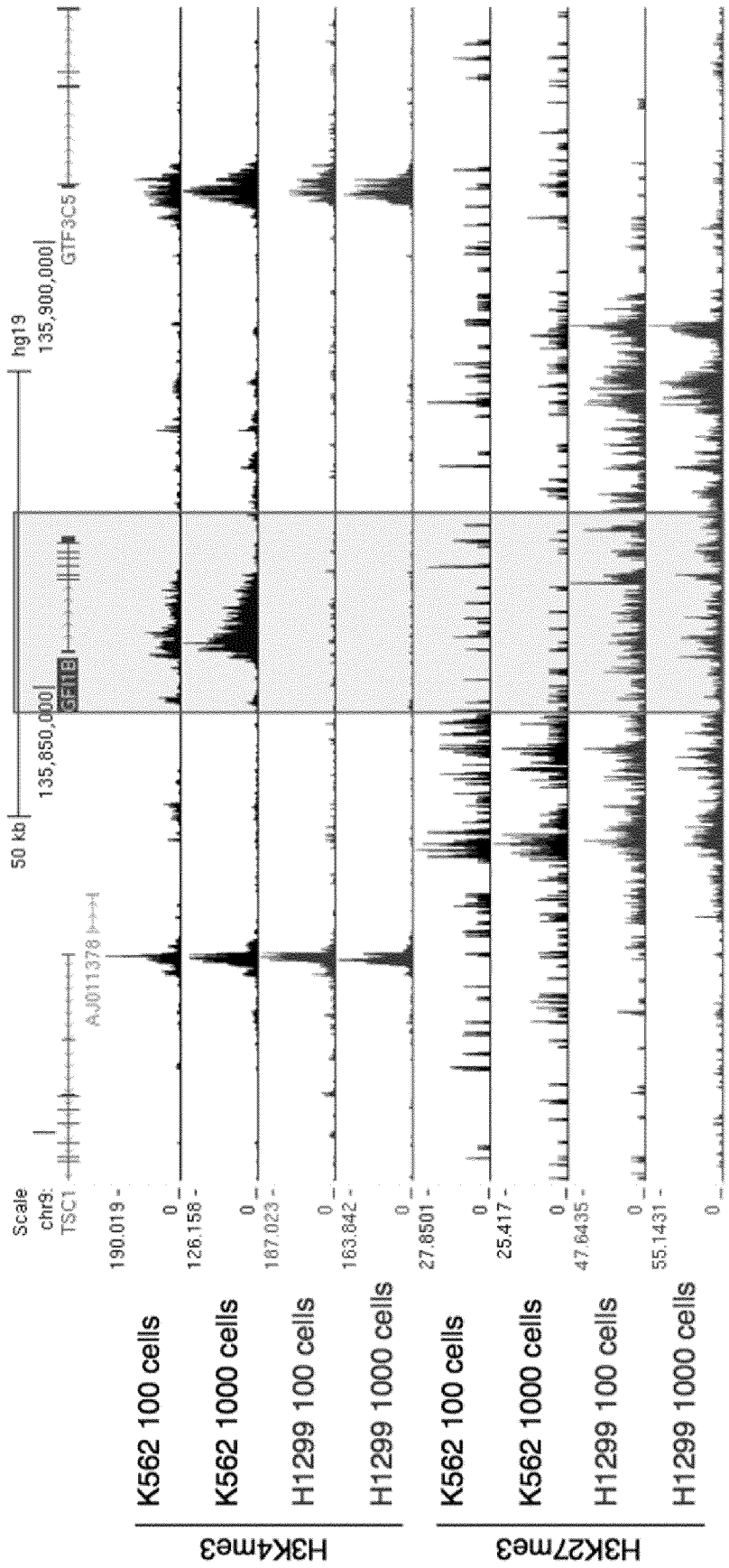

Scale
chr9:
TSC1
AJ011378
50 kb
135,850,000
GFI1B
hg19
135,900,000
GTF3C5
190.019
126.158
187.023
163.842
27.8501
25.417
47.6435
55.1431
0
K562 100 cells
K562 1000 cells
H1299 100 cells
H1299 1000 cells
H3K4me3
H3K27me3

Fig. 6A

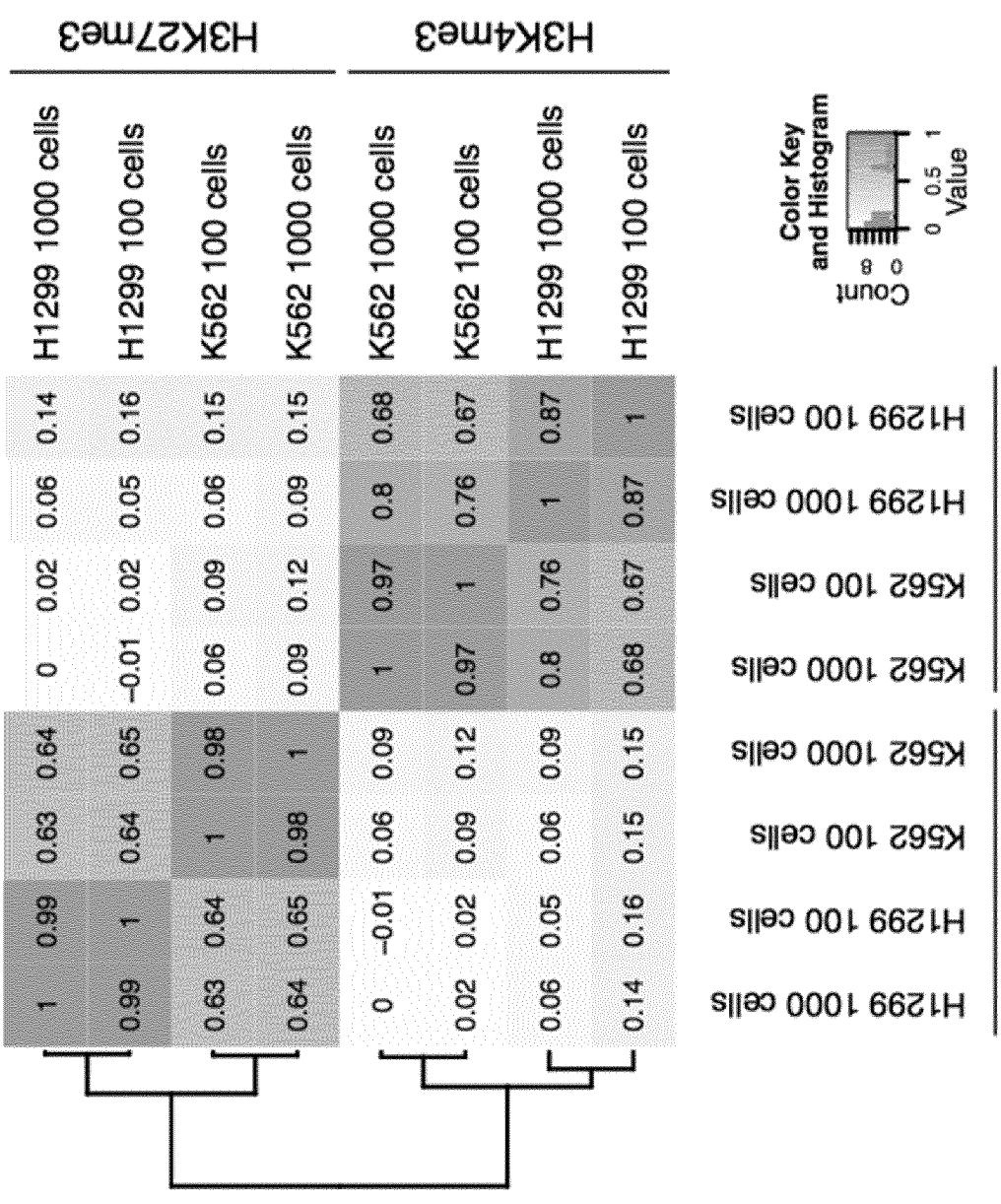

H3K27me3
H3K4me3
H1299 1000 cells
H1299 100 cells
K562 100 cells
K562 1000 cells
H1299 1000 cells
H1299 100 cells
K562 100 cells
K562 1000 cells
Color Key and Histogram
Count
Value
H1299 1000 cells
H1299 100 cells
K562 100 cells
K562 1000 cells
H1299 1000 cells
H1299 100 cells
K562 100 cells
K562 1000 cells

Fig. 6B

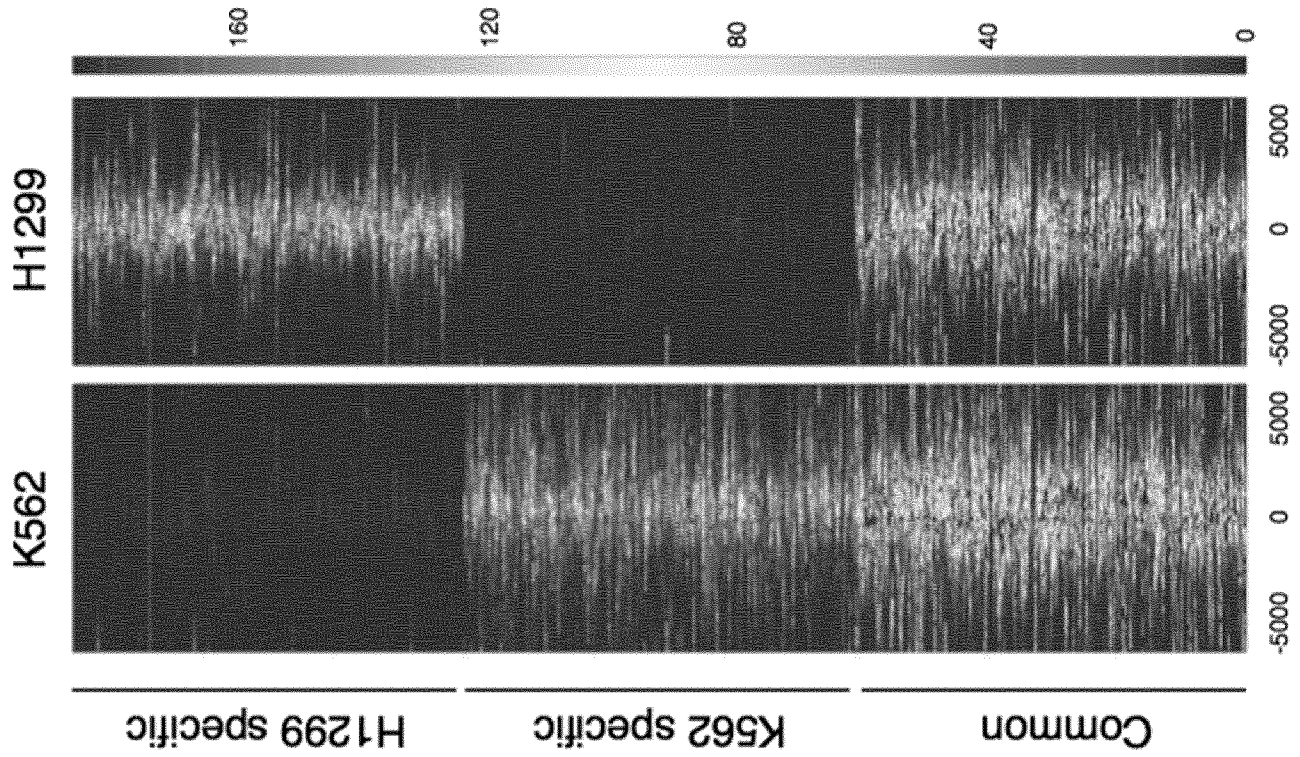

160
120
80
40
0
H1299
K562
5000
0
-5000
5000
0
-5000
H1299 specific
K562 specific
Common

Fig. 6C

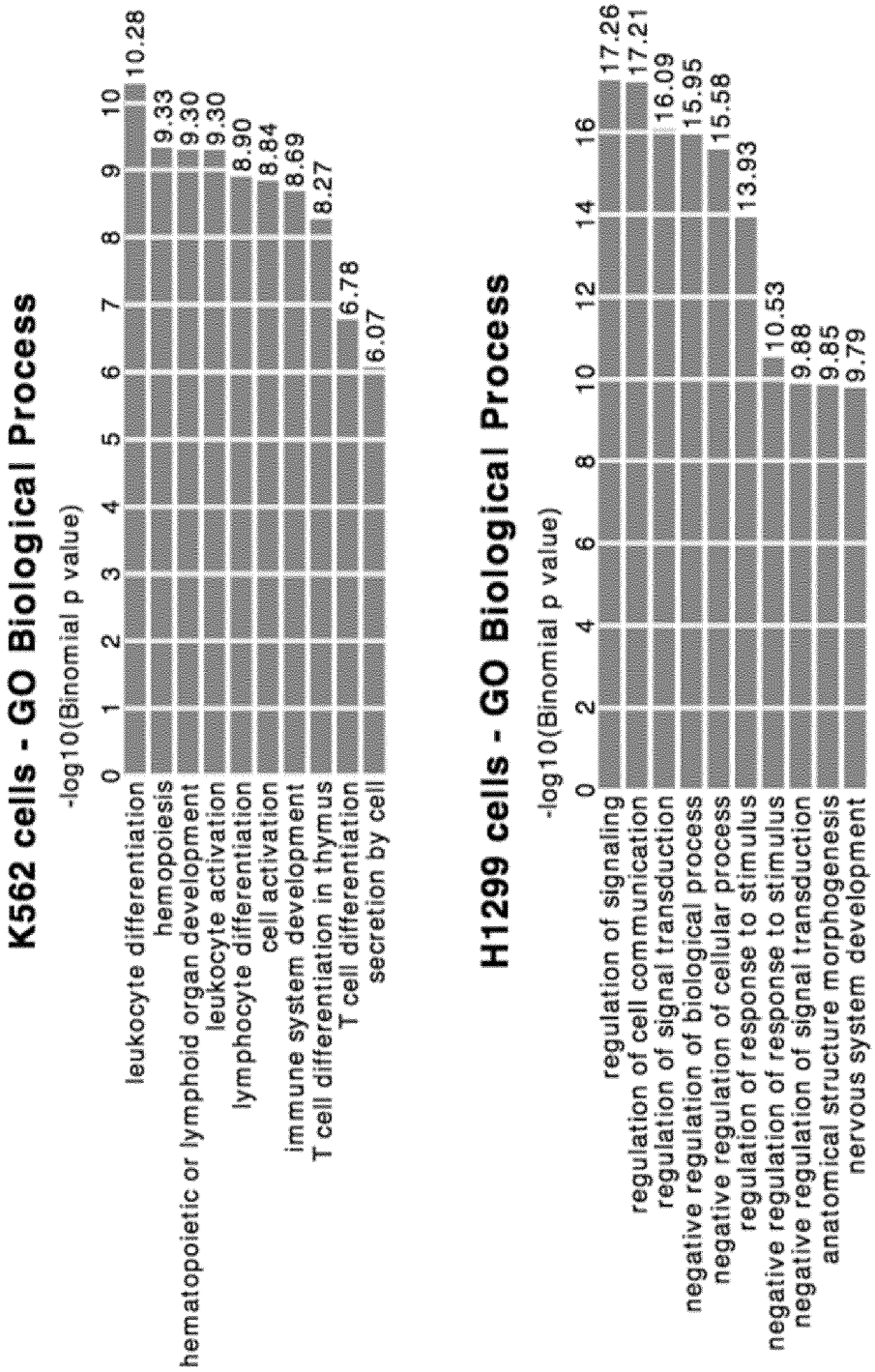

K562 cells - GO Biological Process
-log10(Binomial p value)
0 1 2 3 4 5 6 7 8 9 10
leukocyte differentiation 10.28
hemopoiesis 9.33
hematopoietic or lymphoid organ development 9.30
leukocyte activation 9.30
lymphocyte differentiation 8.90
cell activation 8.84
immune system development 8.69
T cell differentiation in thymus 8.27
T cell differentiation 6.78
secretion by cell 6.07
H1299 cells - GO Biological Process
-log10(Binomial p value)
0 2 4 6 8 10 12 14 16
regulation of signaling 17.26
regulation of cell communication 17.21
regulation of signal transduction 16.09
negative regulation of biological process 15.95
negative regulation of cellular process 15.58
regulation of response to stimulus 13.93
negative regulation of response to stimulus 10.53
negative regulation of signal transduction 9.88
anatomical structure morphogenesis 9.85
nervous system development 9.79

Fig. 6D

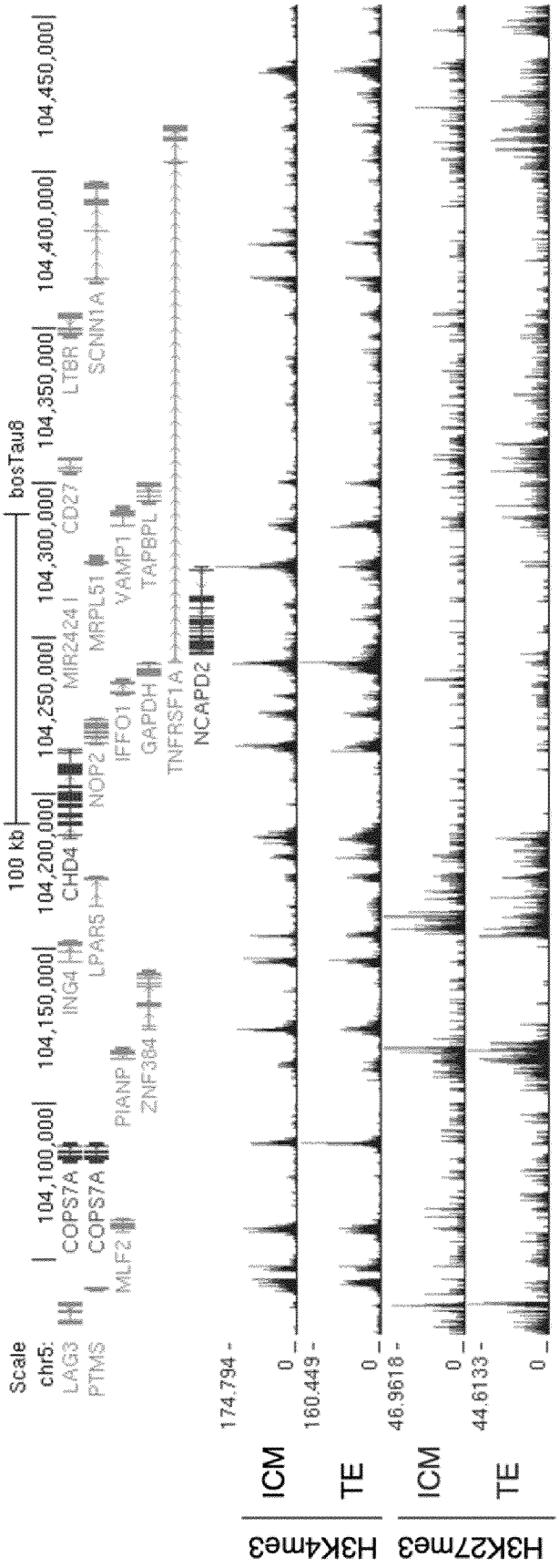
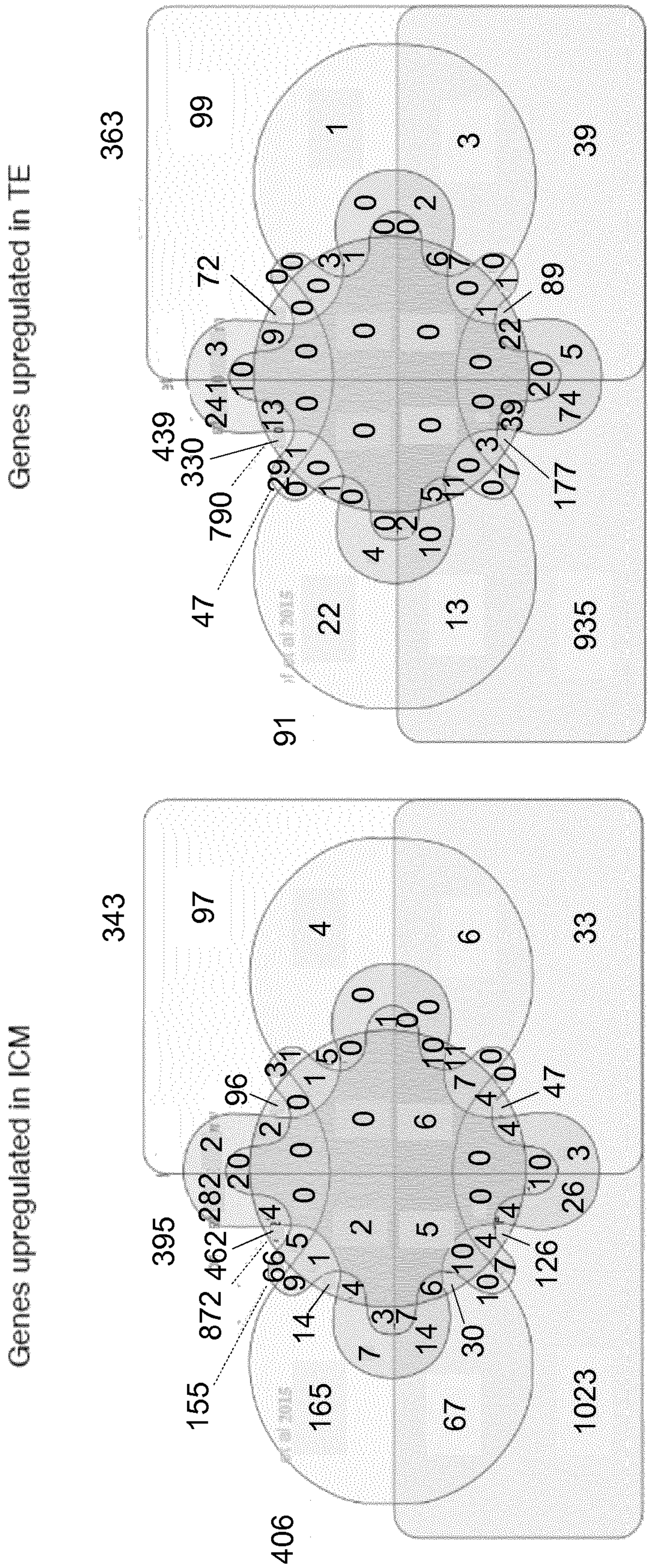

H3K4me3
H3K27me3
ICM
TE
ICM
TE
174.794
160.449
46.9618
44.6133
Scale
chr5:
100 kb
bosTau8
104,100,000
104,150,000
104,200,000
104,250,000
104,300,000
104,350,000
104,400,000
104,450,000
LAG3
PTMS
COPS7A
COPS7A
MLF2
PIANP
ING4
ZNF384
CHD4
LPAR5
NOP2
IFFO1
MIR2424
MRPL51
GAPDH
VAMP1
CD27
TAPBPL
TNFRSF1A
NCAPD2
LTBR
SCNN1A

Fig. 7A



Common genes in nr. of datasets:
2091 - 1 dataset
407 - 2 datasets
136 - 3 datasets
46 - 4 datasets
22 - 5 datasets
6 - 6 datasets

Common genes in nr. of datasets:
1656 - 1 dataset
406 - 2 datasets
175 - 3 datasets
40 - 4 datasets
6 - 5 datasets
0 - 6 datasets

Fig. 7B

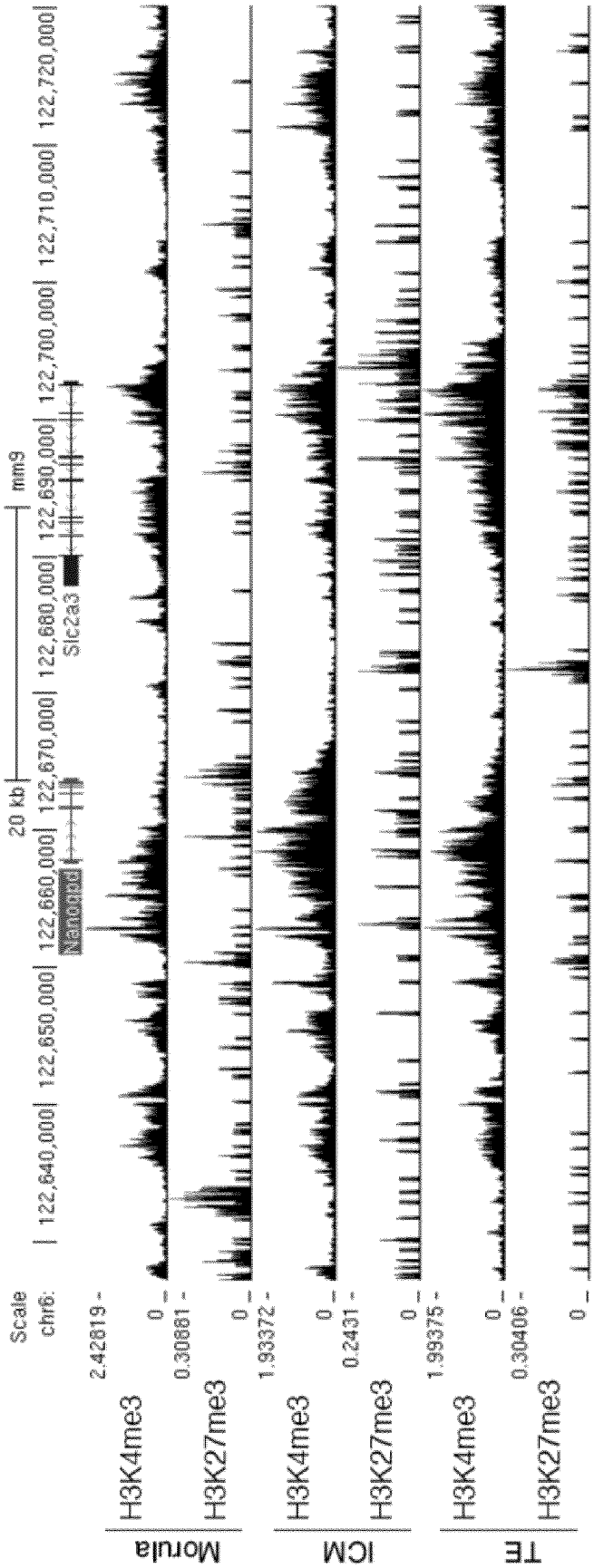

Scale
chr6:
122,640,000
122,650,000
122,660,000
122,670,000
122,680,000
122,690,000
122,700,000
122,710,000
122,720,000
20 kb
mm9
Nanog
Slc2a3
Morula
ICM
TE
H3K4me3
H3K27me3
2.42819
0.30881
1.93372
0.2431
1.99375
0.30406
0

Fig. 7C

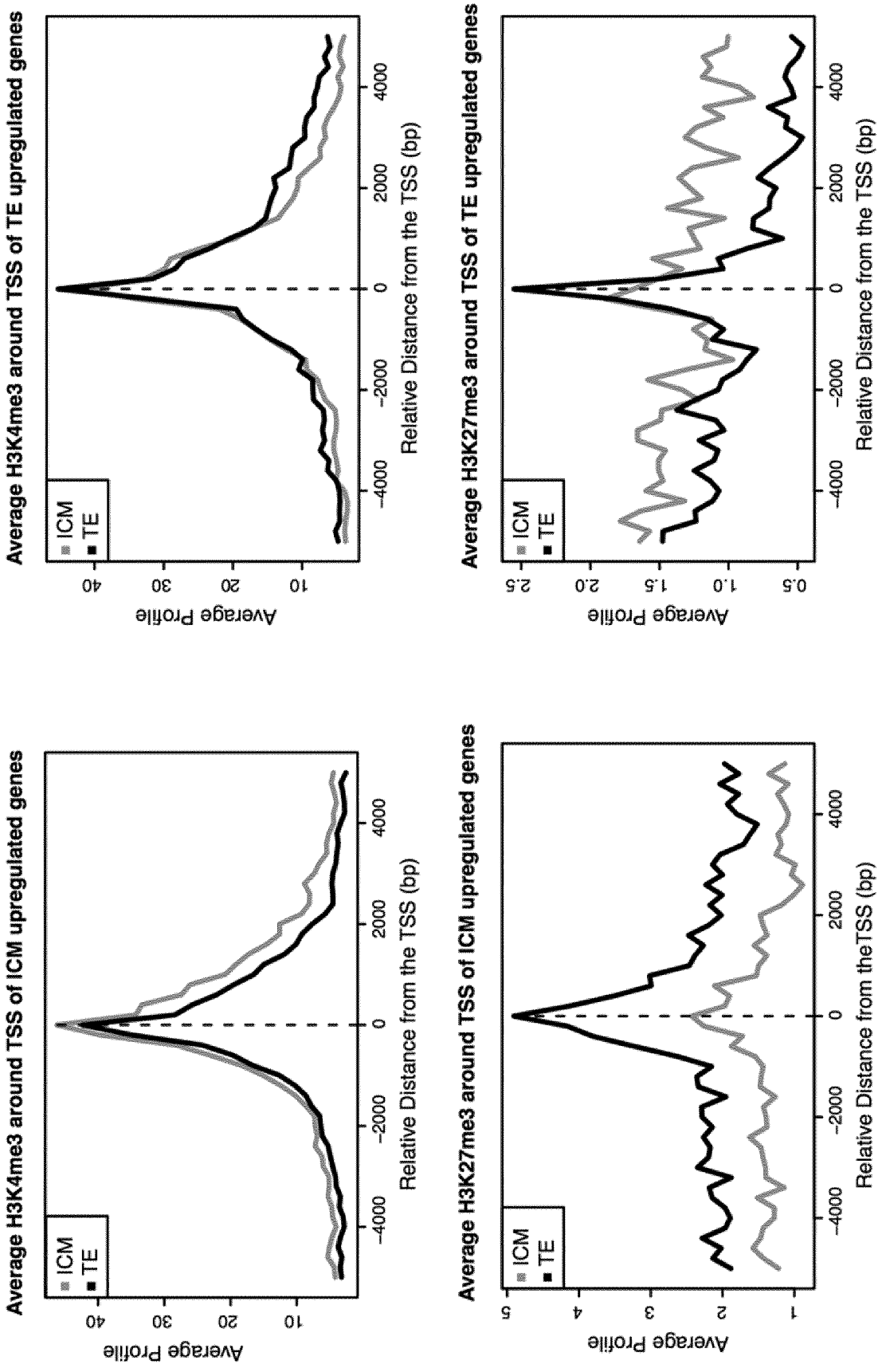

Average H3K4me3 around TSS of ICM upregulated genes
Average H3K4me3 around TSS of TE upregulated genes
Average H3K27me3 around TSS of ICM upregulated genes
Average H3K27me3 around TSS of TE upregulated genes
ICM
TE
Average Profile
Relative Distance from the TSS (bp)
Relative Distance from theTSS (bp)
−4000
−2000
0
2000
4000

Fig. 8A

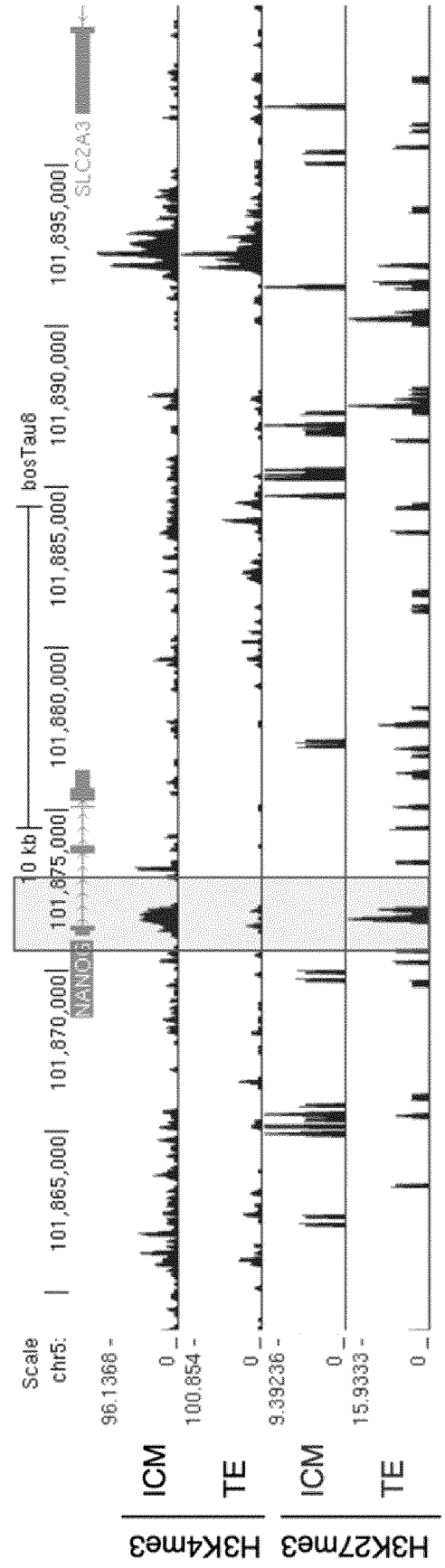

H3K4me3
H3K27me3
ICM
TE
ICM
TE
Scale
chr5:
10 kb
bosTau8
NANOG
SLC2A3
101,865,000
101,870,000
101,875,000
101,880,000
101,885,000
101,890,000
101,895,000
96.1368
100.654
9.39236
15.9333
0

Fig. 8B

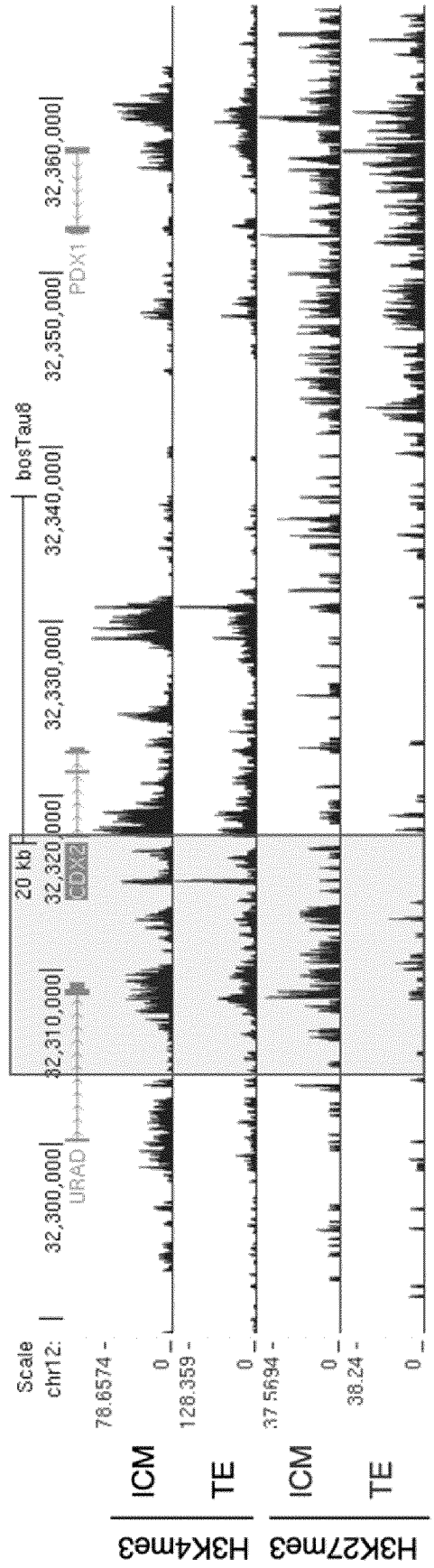

H3K4me3
H3K27me3
ICM
TE
ICM
TE
Scale
chr12:
78.6574
128.359
37.5694
38.24
0
32,300,000
32,310,000
32,320,000
32,330,000
32,340,000
32,350,000
32,360,000
20 kb
bosTau8
URAD
PDX2
PDX1

Fig. 8C

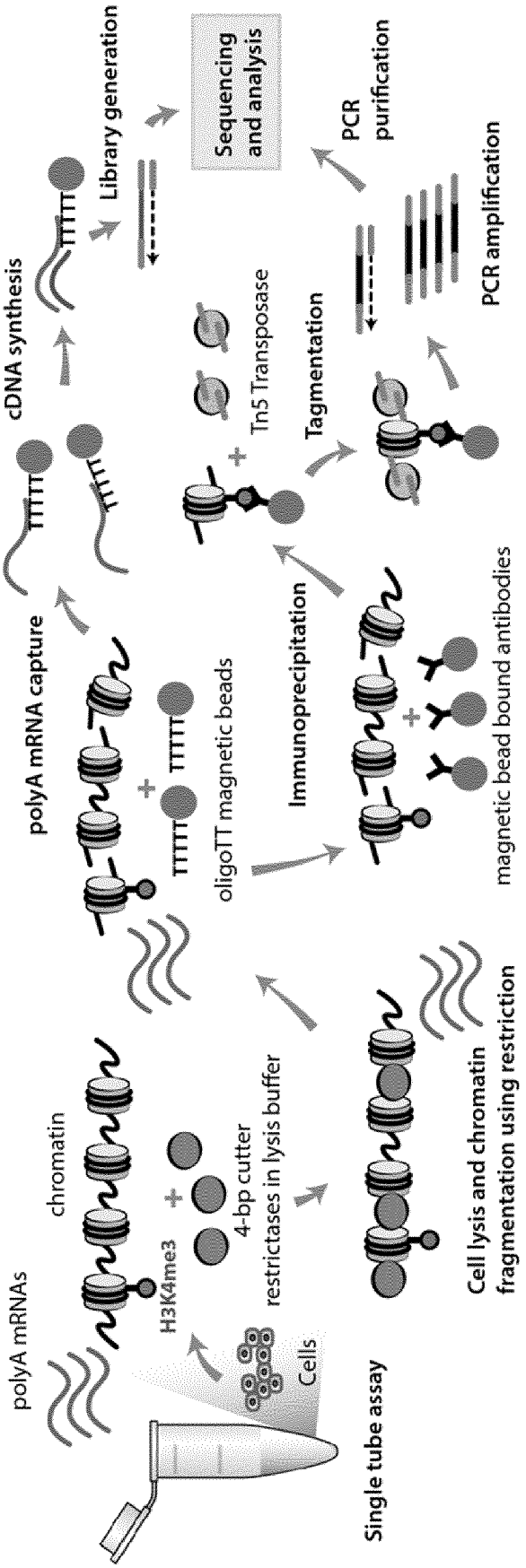

polyA mRNAs
chromatin
H3K4me3
4-bp cutter
restrictases in lysis buffer
Cells
Single tube assay
Cell lysis and chromatin
fragmentation using restriction
polyA mRNA capture
oligoTT magnetic beads
Immunoprecipitation
magnetic bead bound antibodies
Tagmentation
Tn5 Transposase
cDNA synthesis
Library generation
Sequencing
and analysis
PCR
purification
PCR amplification

Fig. 9

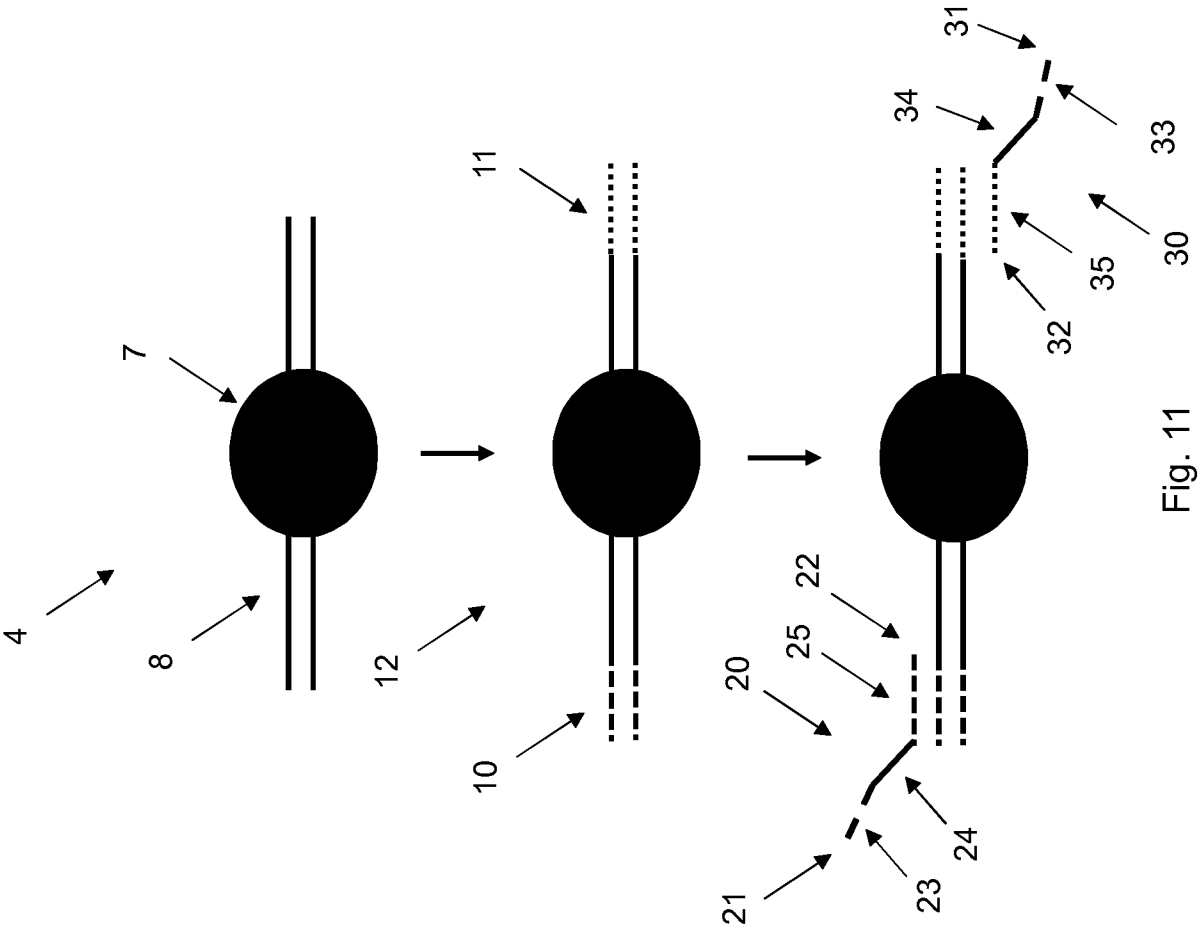

4
8
7
12
10
11
20
25
22
21
23
24
32
35
30
34
31
33

Fig. 11

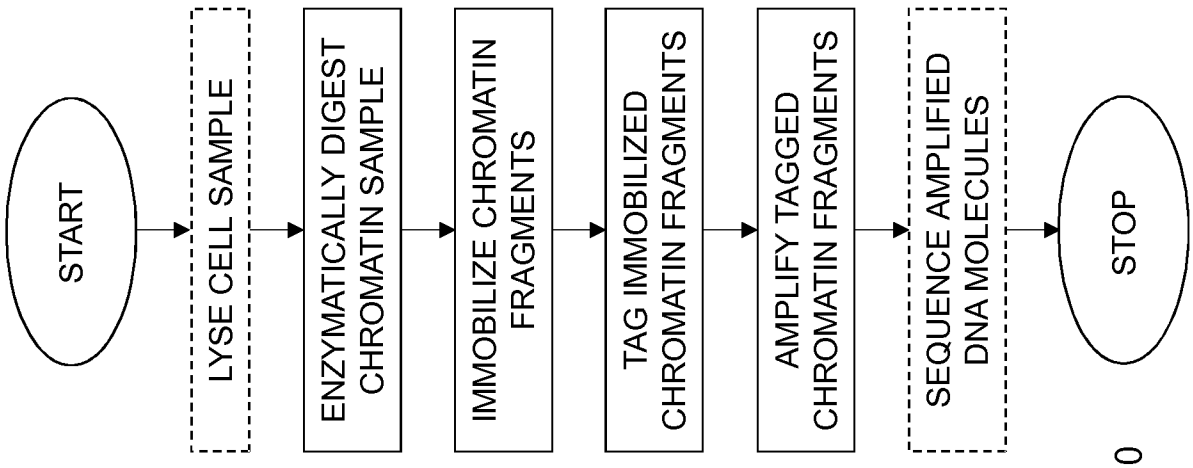

START
LYSE CELL SAMPLE
ENZYMATICALLY DIGEST CHROMATIN SAMPLE
IMMOBILIZE CHROMATIN FRAGMENTS
TAG IMMOBILIZED CHROMATIN FRAGMENTS
AMPLIFY TAGGED CHROMATIN FRAGMENTS
SEQUENCE AMPLIFIED DNA MOLECULES
STOP
S1
S2
S3
S4
S5
S6

Fig. 10

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2018/077934

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/6806 C12Q1/6804
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, BIOSIS, CAB Data, Sequence Search, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2017/025594 A1 (CEMM FORSCHUNGSZENTRUM FÜR MOLEKULARE MEDIZIN GMBH [AT]) 16 February 2017 (2017-02-16) the whole document	1-36
A	-& SCHMIDL C; RENDEIRO AF; SHEFFIELD NC; BOCK C: "ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors (+ Online Methods)", NAT. METHODS, vol. 12, no. 10, October 2015 (2015-10), pages 963-965+2pp, XP002788177, cited in the application the whole document	1-36
X	WO 2014/205296 A1 (BROAD INST INC [US]; HARVARD COLLEGE [US]; GEN HOSPITAL CORP DBA MASSA) 24 December 2014 (2014-12-24) the whole document	24, 27-30, 33,35,36

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
8 February 2019	26/02/2019

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Hornig, Horst

4

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2013/078470 A2 (MOTIF ACTIVE [US]; FERNANDEZ JOSEPH [US]; JELINEK MARY ANNE [US]; EGAN) 30 May 2013 (2013-05-30) the whole document	1-36
A	WO 2014/190214 A1 (ACTIVE MOTIF INC [US]) 27 November 2014 (2014-11-27) the whole document	1-36
A,P	Junaid Akhtar et al.: "TAF-ChIP: An ultra-low input approach for genome wide chromatin immunoprecipitation assay", BioRxiv, 11 April 2018 (2018-04-11), pages 1-20, XP002788178, DOI: 10.1101/299727 Retrieved from the Internet: URL:https://www.biorxiv.org/content/early/2018/04/11/299727.abstract [retrieved on 2019-01-17] the whole document	1-36
A	WO 2017/034970 A1 (MASSACHUSETTS GEN HOSPITAL [US]; SEQLL LLC [US]; SHEMA-YAACOBY EFRAT [) 2 March 2017 (2017-03-02) the whole document	1-36
A	-& SHEMA EFRAT ET AL: "Single-molecule decoding of combinatorially modified nucleosomes", SCIENCE, vol. 352, no. 6286, 6 May 2016 (2016-05-06), pages 717-721, XP002764306, , AAAS, AMERICAN ASSOC. FOR THE ADVANCEMENT OF SCIENCE, US ISSN: 0036-8075 the whole document	1-36
A	MEYER CLIFFORD A ET AL: "Identifying and mitigating bias in next-generation sequencing methods for chromatin biology.", NATURE REVIEWS GENETICS, vol. 15, no. 11, November 2014 (2014-11), pages 709-721, XP002788179, ISSN: 1471-0064, DOI: 10.1038/nrg3788 the whole document	1-36
A	WO 2011/096926 A1 (MASSACHUSETTS INST TECHNOLOGY [US]; BROAD INST INC [US]; ADLI MAZHAR [) 11 August 2011 (2011-08-11) cited in the application the whole document	1-36

-/--

4

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JASON D BUENROSTRO ET AL: "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position", NATURE METHODS , vol. 10, no. 12 1 December 2013 (2013-12-01), pages 1213-1218, XP002764310, NATURE PUBLISHING GROUP, GB ISSN: 1548-7105, DOI: 10.1038/NMETH.2688 Retrieved from the Internet: URL:http://www.nature.com/nmeth/journal/v10/n12/pdf/nmeth.2688.pdf [retrieved on 2013-10-06] the whole document	1-36
A	KIDDER: "'ChIP-Seq: technical considerations for obtaining high-quality data.'", NATURE IMMUNOLOGY, vol. 12, no. 10, 1 October 2011 (2011-10-01), pages 918-922, XP055240604, DOI: 10.1038/ni.2117 the whole document	1-36
A	PARK PETER J: "ChIP-seq: advantages and challenges of a maturing technology", NATURE REVIEWS GENETICS, vol. 10, no. 10, 1 October 2009 (2009-10-01), pages 669-680, XP002604157, NATURE PUBLISHING GROUP, GB ISSN: 1471-0056, DOI: 10.1038/NRG2641 [retrieved on 2009-09-08] the whole document	1-36
A	"ChIP-IT Express Magnetic Chromatin Immunoprecipitation Kits (version D2)", INTERNET CITATION, 2009, pages 1-36, XP002755832, Retrieved from the Internet: URL:http://www.biotechniques.com/multimedia/archive/00054/chip-it_express_manu_54059a.pdf [retrieved on 2016-03-21] the whole document	1-36
A	WO 2010/084605 A1 (NEC DISPLAY SOLUTIONS LTD [JP]; WAKASUGI SATOSHI [JP]) 29 July 2010 (2010-07-29) the whole document	1-36

-/--

4

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2016/061517 A2 (ILLUMINA CAMBRIDGE LTD [GB]; STEEMERS FRANK J [US]; GUNDERSON KEVIN L) 21 April 2016 (2016-04-21) the whole document	1-36

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2018/077934

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2017025594 A1	16-02-2017	CA 2995305 A1	16-02-2017
		CN 108368540 A	03-08-2018
		EP 3334841 A1	20-06-2018
		JP 2018532373 A	08-11-2018
		US 2018237951 A1	23-08-2018
		WO 2017025594 A1	16-02-2017
WO 2014205296 A1	24-12-2014	US 2016208323 A1	21-07-2016
		WO 2014205296 A1	24-12-2014
WO 2013078470 A2	30-05-2013	NONE	
WO 2014190214 A1	27-11-2014	EP 2999784 A1	30-03-2016
		HK 1221976 A1	16-06-2017
		US 2016115474 A1	28-04-2016
		WO 2014190214 A1	27-11-2014
WO 2017034970 A1	02-03-2017	NONE	
WO 2011096926 A1	11-08-2011	US 2011189677 A1	04-08-2011
		WO 2011096926 A1	11-08-2011
WO 2010084605 A1	29-07-2010	JP 5327904 B2	30-10-2013
		JP WO2010084605 A1	12-07-2012
		WO 2010084605 A1	29-07-2010
WO 2016061517 A2	21-04-2016	AU 2015331739 A1	11-05-2017
		BR 112017007912 A2	23-01-2018
		CA 2964799 A1	21-04-2016
		JP 2018501776 A	25-01-2018
		KR 20170107423 A	25-09-2017
		RU 2017116989 A	19-11-2018
		SG 11201703139V A	28-07-2017
		US 2019040382 A1	07-02-2019
		US 2019048332 A1	14-02-2019
		WO 2016061517 A2	21-04-2016