NGS library construction using fragmented/size selected DNA

There are several important considerations when preparing libraries from DNA samples.

the amount of starting material
whether the application is for resequencing or de novo sequencing

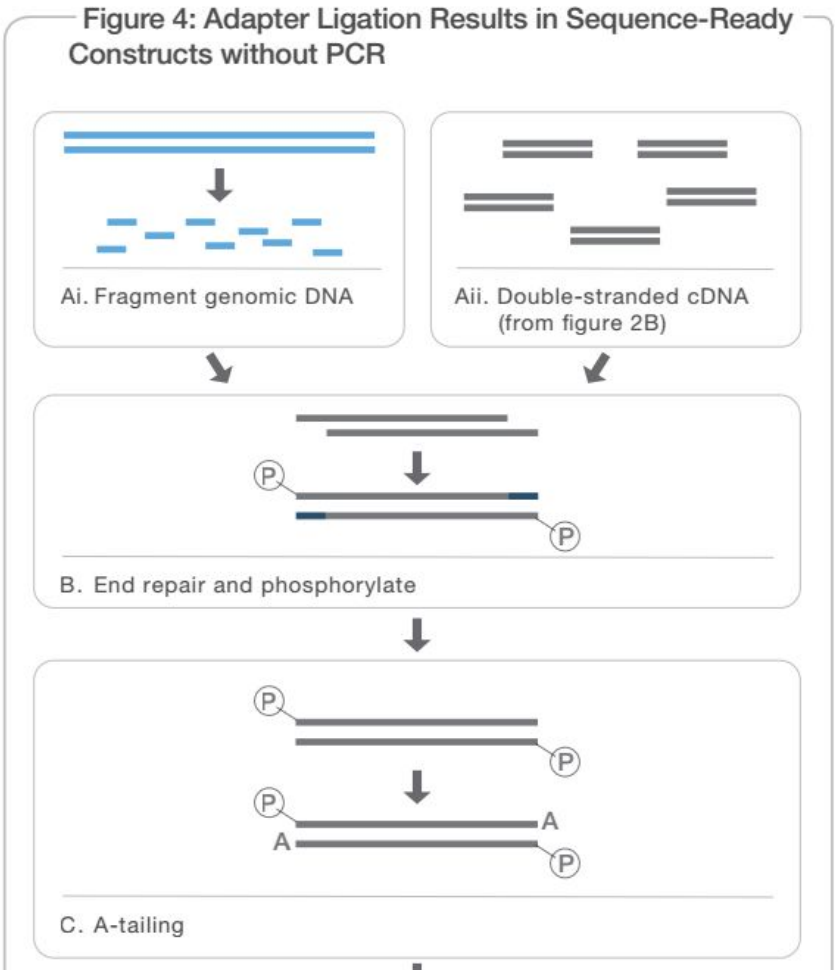
- Library preparations can be susceptible to bias resulting from genomes that contain unusually **high or low GC content**
- approaches have been developed to address these situations through careful selection of polymerases for PCR amplification, thermocycling, conditions and buffers

- Library preparation from DNA samples for sequencing
  - whole genomes,
  - targeted regions within genomes (for example exome sequencing),
  - ChIP-seq experiments,
  - or PCR amplicons follows the same general workflow.
- Ultimately, for any application, the goal is to make the libraries as complex as possible

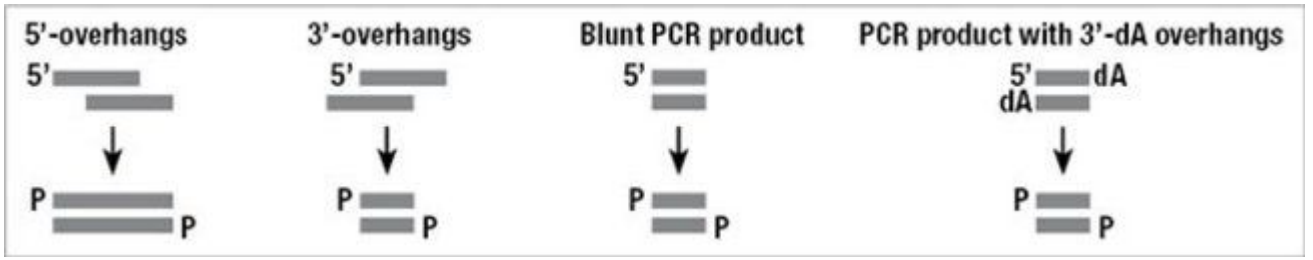

PARS-SEQ
SHAPE-SEQ
MDA
MALBAC
CHIRP-SEQ
RNA-SEQ
MAINE-SEQ
RRBS-SEQ
RIBO-SEQ
ICE
OS-SEQ
CLASH-SEQ
TIF-SEQ/PEAT
IN-SEQ
TAB-SEQ
CHIP-SEQ
5-C
SMMIP
OXBS-SEQ
RIP-SEQ
4-C
TC-SEQ
NET-SEQ
UMI
CAP-SEQ
FAIRE-SEQ
DUPLEX-SEQ
DNASE-SEQ
PAR-CLIP-SEQ
BS-SEQ
MEDIP-SEQ
GRO-SEQ
MERIP-SEQ
DIGITAL
CHIA-PET
ATAC-SEQ
CIP-TAP
ICLIP
HITS-CLIP
MBDCAP-SEQ
PARE-SEQ/GMUT
HI-C/3-C
TRAP-SEQ
RC-SEQ
FRAG-SEQ

- Numerous kits for making sequencing libraries from DNA are available commercially from a variety of vendors.
- Kits are available for making libraries from **microgram down to picogram** quantities of starting material.
- However, one should keep in mind the general principle that more starting material means less amplification and thus better library complexity.

TruSeq DNA Sample Preparation

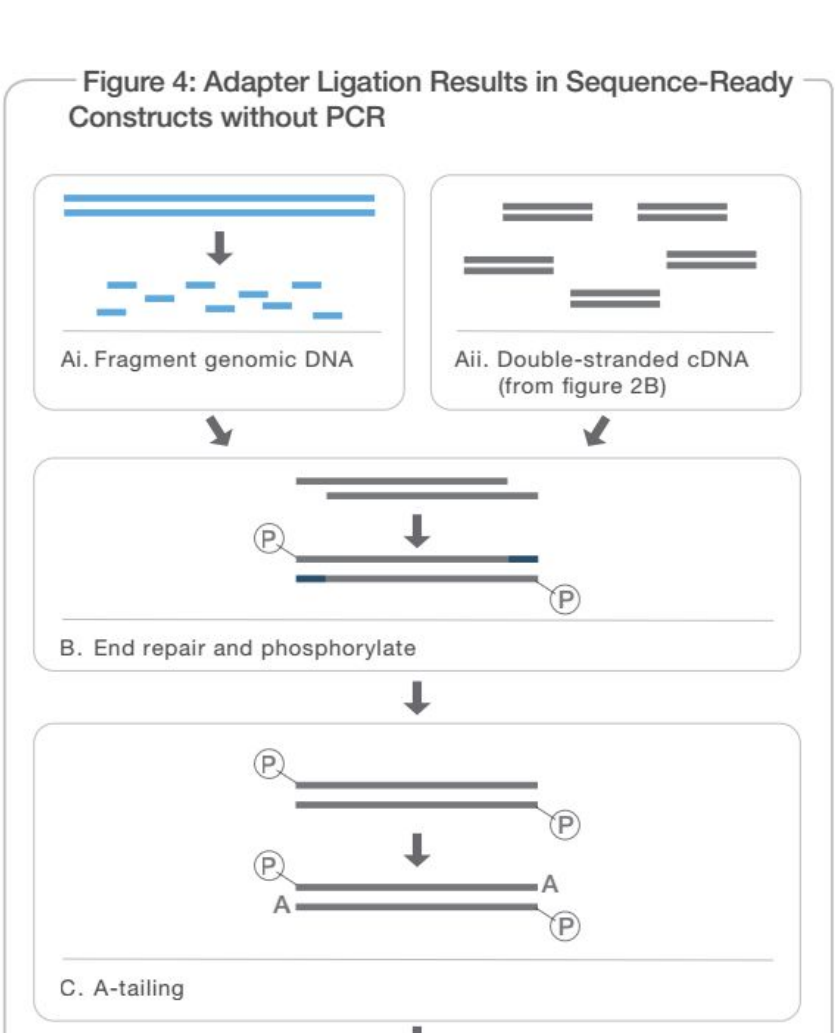


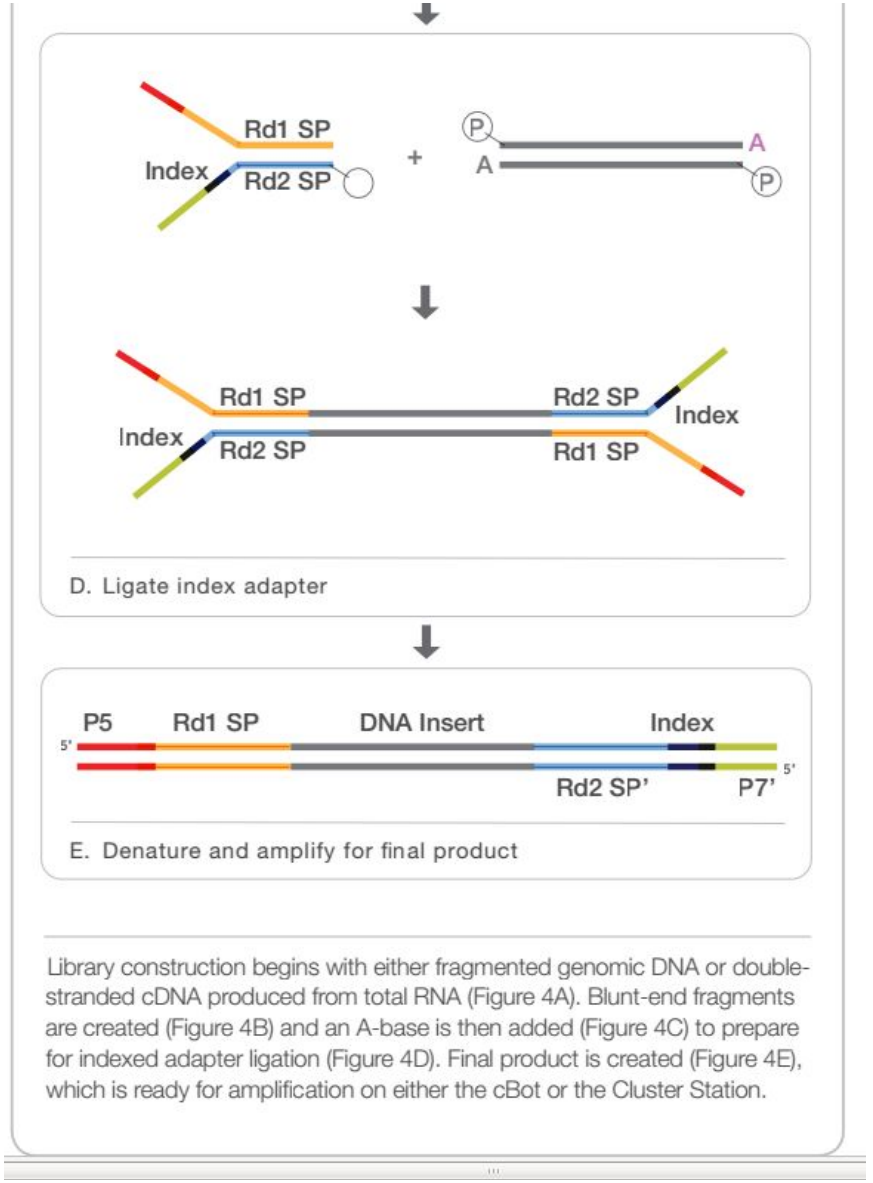
The End Repair Enzyme Mix contains an optimized mixture of **T4 DNA Polymerase** and **Klenow Fragment** to achieve highly effective blunting of fragmented DNA, and **T4 Polynucleotide Kinase (PNK)** for efficient phosphorylation of DNA ends.



- During the DNA end repair reaction, fragmented DNA is converted into blunt-end DNA containing a 5'-phosphate and 3'-hydroxyl groups.
- The 5'→3' polymerase activity of the End Repair Enzyme Mix fills-in 5' protruded DNA ends while 3'→5' exonuclease activity removes 3'-overhangs.
- T4 PNK adds 5'-phosphates to ends of unphosphorylated DNA fragments, such as PCR products.

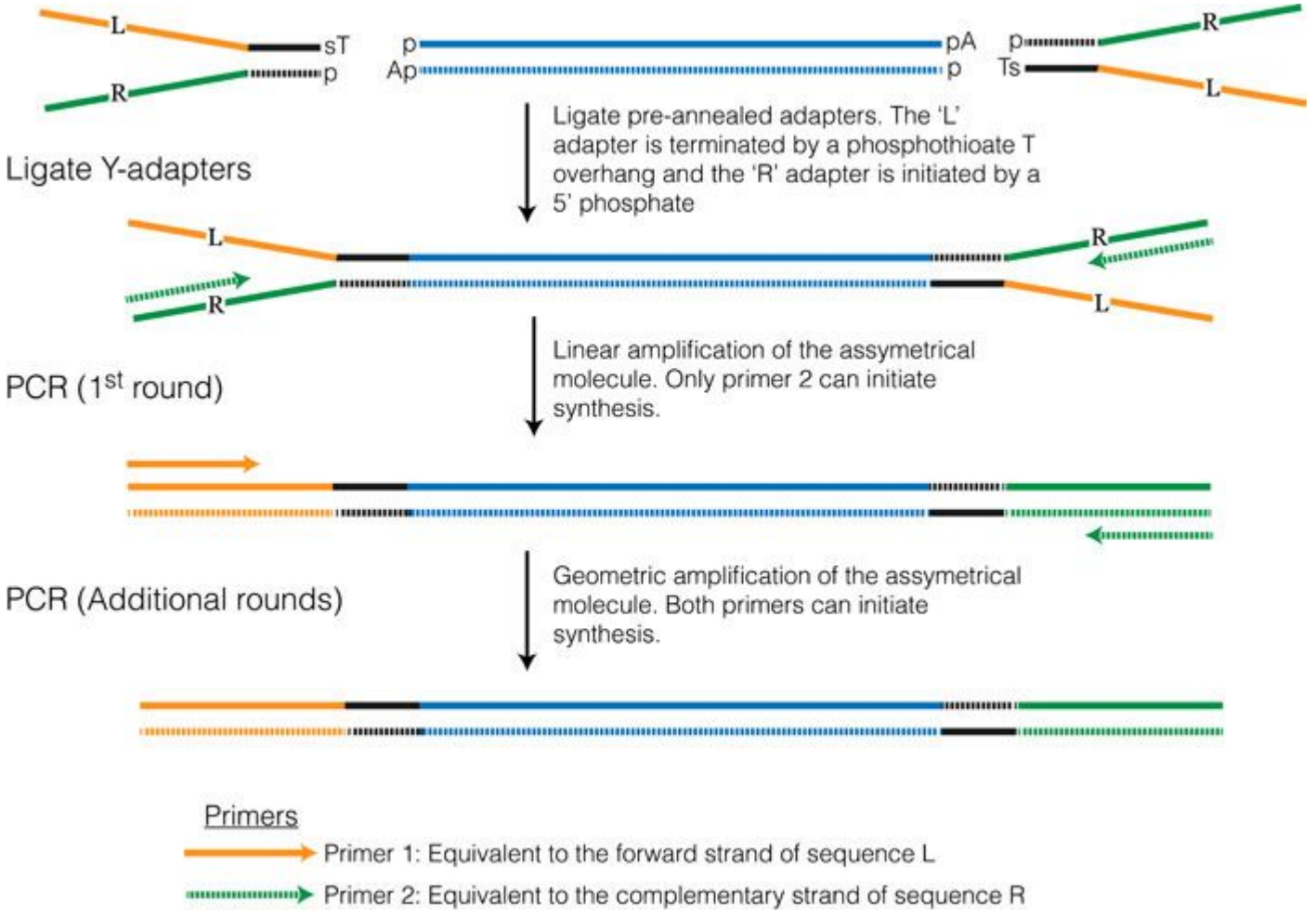
TruSeq DNA Sample Preparation





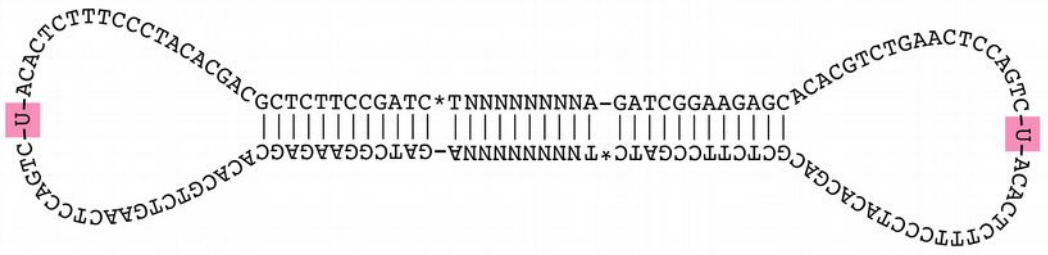
Library construction begins with either fragmented genomic DNA or double-stranded cDNA produced from total RNA (Figure 4A). Blunt-end fragments are created (Figure 4B) and an A-base is then added (Figure 4C) to prepare for indexed adapter ligation (Figure 4D). Final product is created (Figure 4E), which is ready for amplification on either the cBot or the Cluster Station.

Y-adapters illumina

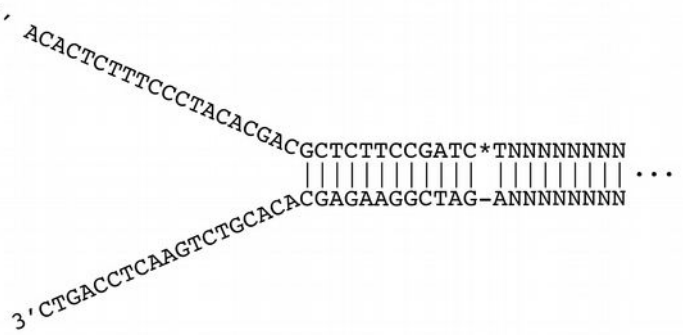


Adapters NEB

1

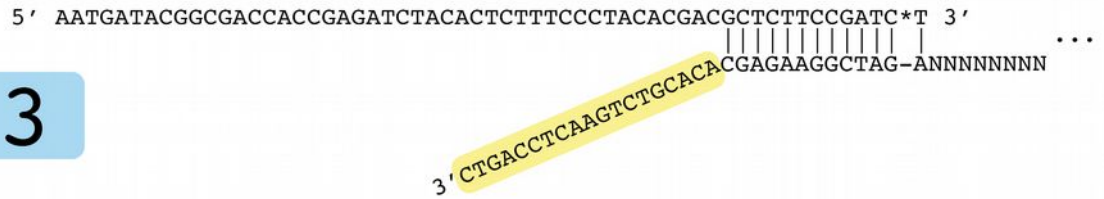


2



USER (Uracil-Specific Excision Reagent) Enzyme generates a single nucleotide gap at the location of a uracil residue.

3



TruSeq® DNA PCR-Free Sample Preparation Kit

DNA Input Recommendations

For best results, follow the input recommendations. Quantify the input gDNA and assess the gDNA quality before beginning library preparation.

- For a 350 bp insert size, use 1 µg input gDNA.
- For a 550 bp insert size, use 2 µg input gDNA.
- Input amounts lower than those specified results in low yield and increased duplicates.

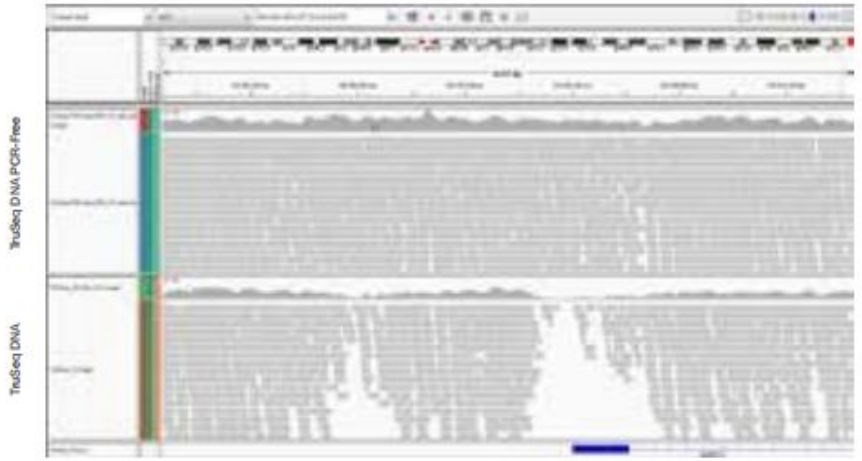
Quantify Input DNA

Use the following recommendations to quantify input DNA:

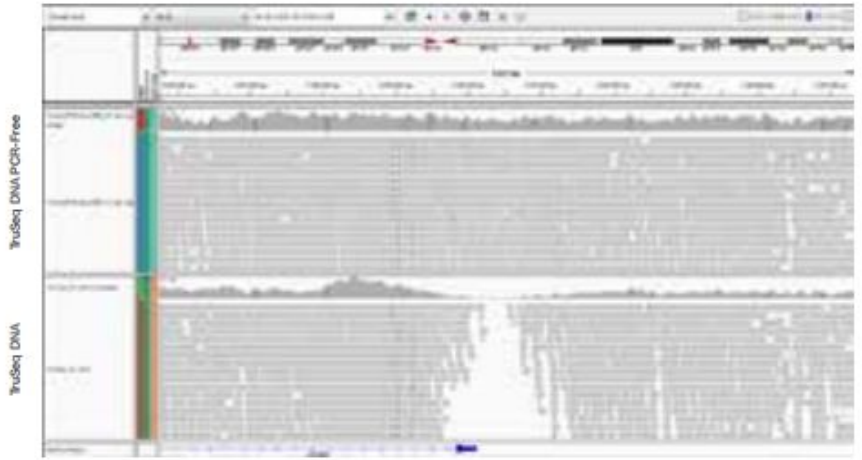
- Successful library preparation depends on accurate quantification of input DNA. To verify results, use multiple methods.
- Use fluorometric-based methods for quantification, such as Qubit or PicoGreen.
- DNA quantification methods that rely on intercalating fluorescent dyes measure only double-stranded DNA and are less subject to the presence of excess nucleic acids.
- Do not use spectrophotometric-based methods, such as NanoDrop, which measure the presence of nucleotides and can result in an inaccurate measurement of gDNA.
- Quantification methods depend on accurate pipetting methods. Do not use pipettes at the extremes of volume specifications. Make sure that pipettes are calibrated.

Figure 6: PCR-Free Protocol Eliminates Coverage Gaps in GC-Rich Content

A



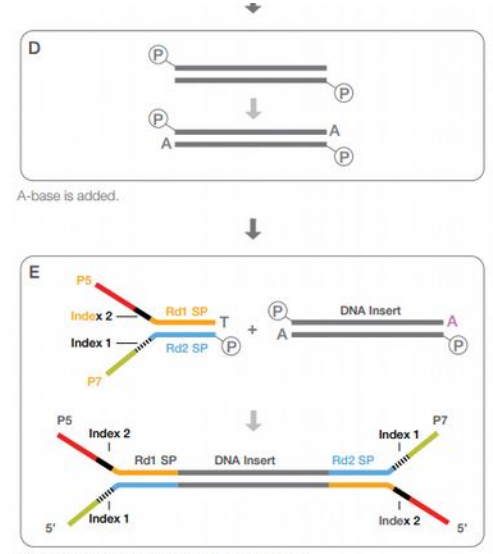
B

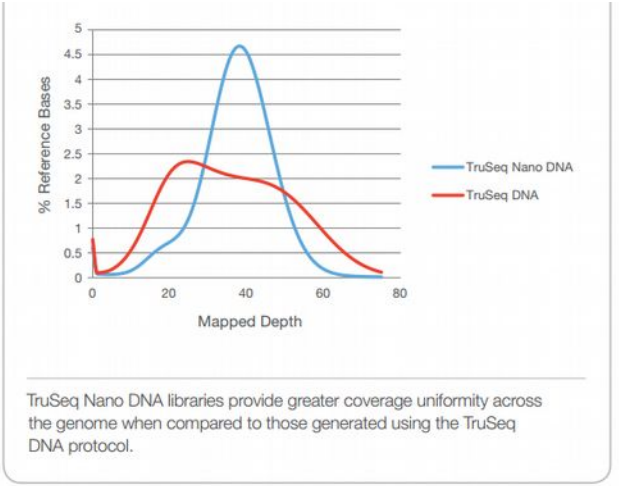


Increased coverage of TruSeq DNA PCR-Free libraries results in fewer coverage gaps, demonstrated here in the GC-rich coding regions of the *RNPEPL1* promoter (A) and the *CREBBP* promoter (B). PCR-Free sequence information is shown in the top panels of A and B, while sequence data generated using TruSeq DNA protocol (with PCR) are shown in the lower panels.

TruSeq® Nano DNA Sample

- The TruSeq Nano DNA protocol supports shearing by Covaris ultrasonication, requiring 100 ng of input DNA for an average insert size of 350 bp or 200 ng DNA for an average insert size of 550 bp.





he TruSeq Nano DNA LT indexing solution features a single-index adapter
Step E.

DNA in a Human cell

If you know the number of base pairs in the genome of your cell of interest, which is about 3 billion in the haploid human genome, you may calculate as follows:

3x10^9 bp x 2 (diploid) x 660 (AVGed MW of 1 bp) x 1.67x10^-12pg ("weight in dalton") =

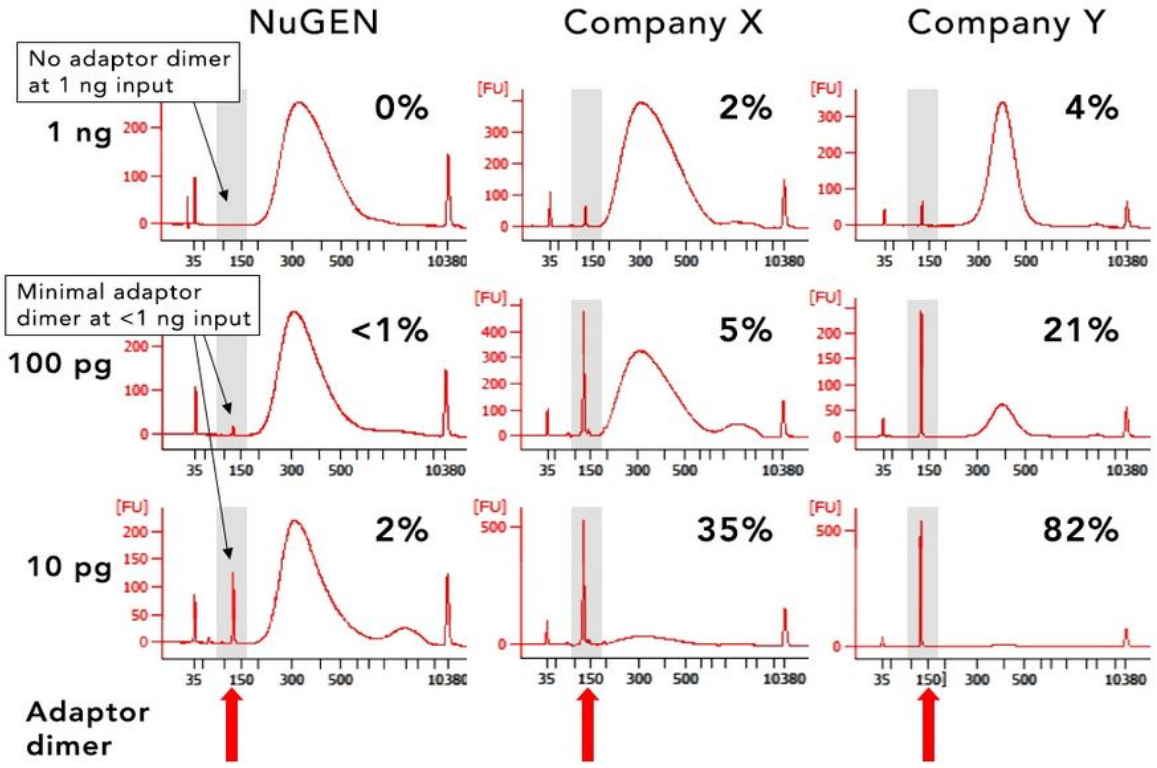
6.6pg/ diploid primary cell.

$$1\,u \simeq 1,66 \times 10^{-24}\,g$$

PureGenome™ Low Input NGS Library Construction Kit

Illumina NGS prep from as little as 50 pg of input DNA

The PureGenome™ Low Input NGS Library Construction Kit streamlines the process of generating indexed or barcoded libraries for Illumina® HiSeq® sequencing from extremely low amounts of input DNA (from 50 pg to 1 ng). This kit is well suited for applications such as ChIP-Seq (chromatin immunoprecipitation followed by next generation sequencing) as well as other sequencing applications where input DNA is limited.

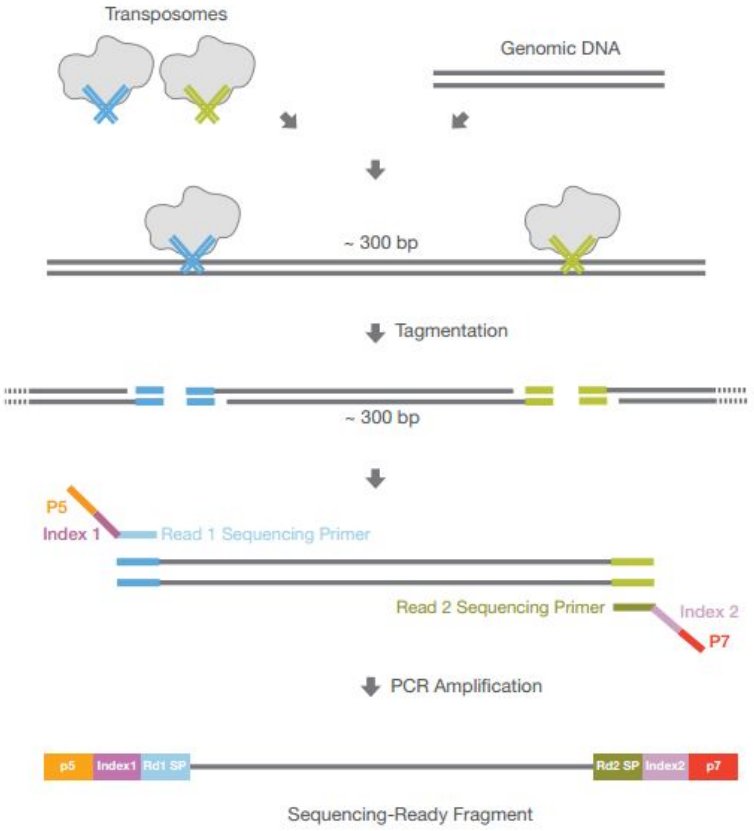


The Ovation Ultralow System V2 produces significantly less adaptor dimers than competing kits. The lack of adaptor dimers enables library construction even at very low inputs making this system ideal for a range of inputs from 100 ng down to 10 pg.

Nextera®
DNA Sample Preparation Kits

- Easiest to Use
  - Prepare sequencing-ready samples in 1.5 hours with 15 minutes hands-on time
- Lowest DNA Input
  - Use just 50 ng DNA per sample, enabling use with samples in limited supply

Figure 2: Nextera Sample Preparation Biochemistry



Nextera chemistry simultaneously fragments and tags DNA in a single step. A simple PCR amplification then appends sequencing adapters and sample indexes to each fragment.

Table 2: Representative Nextera Applications

Examples of Nextera Applications
Large-genome resequencing
Small-genome resequencing
Amplicon resequencing
Clone or plasmid sequencing

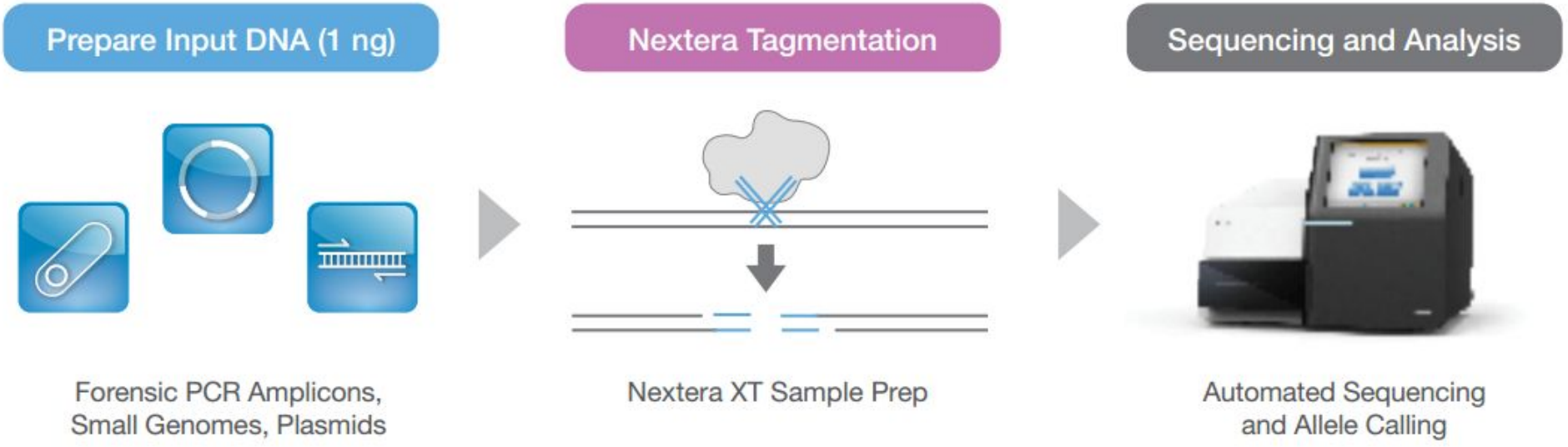
References

1. Ramirez MS, Adams MD, Bonomo RA, Centrón D, et al. (2011) Genomic analysis of *Acinetobacter baumannii* A118 by comparison of optical maps: Identification of structures related to its susceptibility phenotype. Antimicrob Agents Chemother, 55(4): 1520–6.
2. Adey A, Morrison HG, Asan, Xun X, Kitzman JO, et al. (2010) Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. Genome Biol 11: R119.
3. Bimber BN, Dudley DM, Lauck M, Becker EA, Chin EN, et al. (2010) Whole-genome characterization of human and simian immunodeficiency virus intrahost diversity by ultradeep pyrosequencing. J Virol 84: 12087–92.
4. Kitzman JO, Mackenzie AP, Adey A, Hiatt JB, Patwardhan RP, et al. (2010) Haplotype-resolved genome sequencing of a Gujarati Indian individual. Nat Biotechnol 29: 59–63.
5. Linnarsson, S. (2010) Recent advances in DNA sequencing methods - General principles of sample preparation. Exp Cell Res 316: 1339–43.
6. Sudmant PH, Kitzman JO, Antonacci F, Alkan C, Malig M, et al. (2010) Diversity of human copy number variation and multicopy genes. Science 330: 641–646.
7. Voelkerding KV, Dames S, and JD Durtschi (2010) Next generation sequencing for clinical diagnostics-Principles and application to targeted resequencing for hypertrophic cardiomyopathy. J Mol Diagn 12: 539–551.

Accelerated Applications

Nextera XT

Figure 1: Nextera XT Sample Preparation Workflow



The combination of Nextera XT and rapid sequencing with the MiSeq System provides a complete DNA to data workflow in only 8 hours.

Focused investigation of key genes

With targeted resequencing, a subset of genes or regions of the genome are isolated and sequenced.

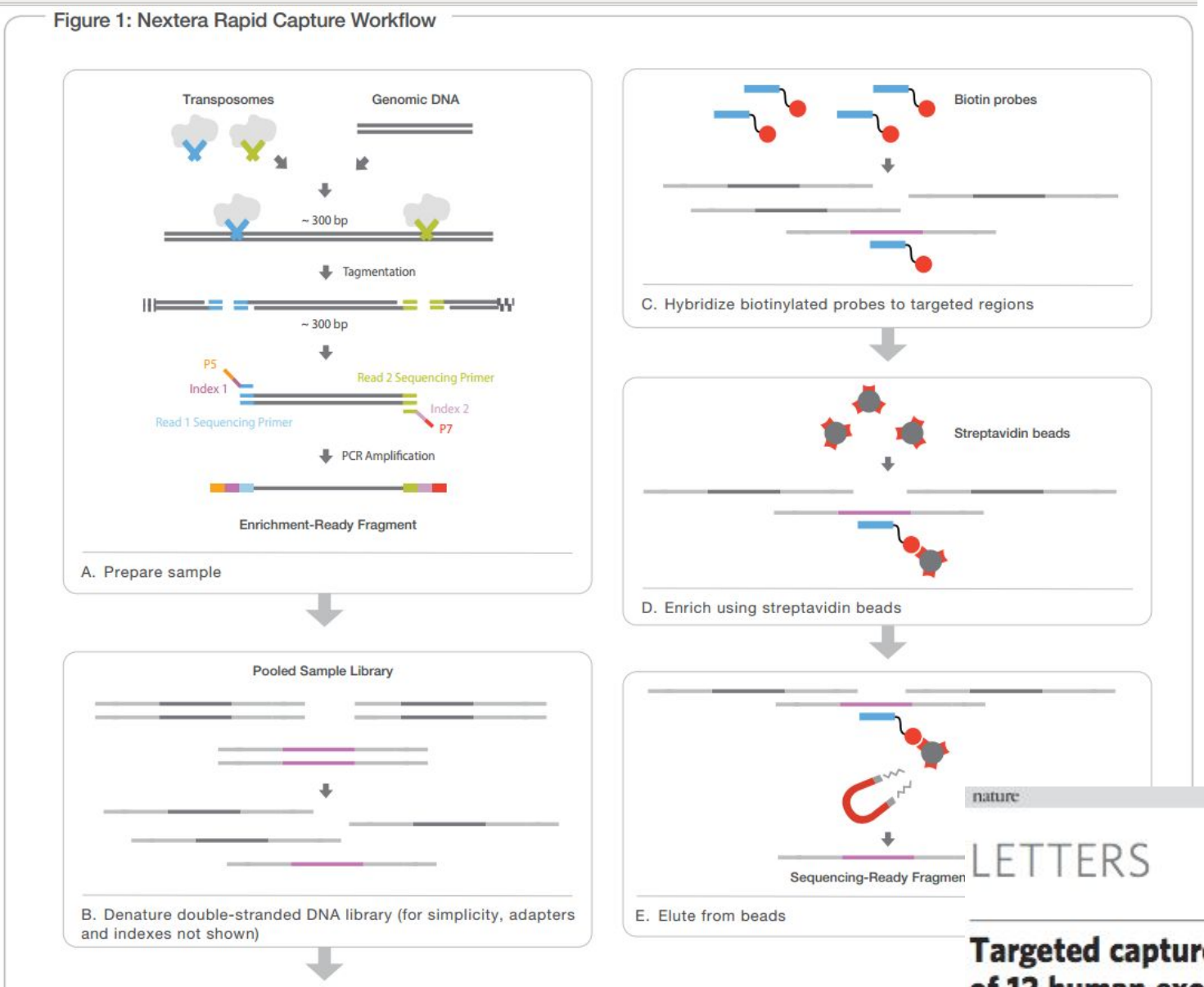
Targeted approaches using next-generation sequencing (NGS) allow researchers to focus time, expenses, and data analysis on specific areas of interest.

Such targeted analysis can include the exome (the protein-coding portion of the genome), specific genes of interest (custom content), targets within genes, or mitochondrial DNA.

Target enrichment:

Regions of interest are captured by hybridization to biotinylated probes and then isolated by magnetic pulldown. Target enrichment captures 20 kb–62 Mb regions, depending on the experimental design.

Rapid Capture Exomes



nature Vol 461|10 September 2009 doi:10.1038/nature08250

LETTERS

Targeted capture and massively parallel sequencing of 12 human exomes

Sarah B. Ng[1], Emily H. Turner[1], Peggy D. Robertson[1], Steven D. Flygare[1], Abigail W. Bigham[2], Choli Lee[1], Tristan Shaffer[1], Michelle Wong[1], Arindam Bhattacharjee[4], Evan E. Eichler[1,3], Michael Bamshad[2], Deborah A. Nickerson[1] & Jay Shendure[1]

Target Capture Workflow

= target sequences
= off-target sequences
= adapter sequences
= biotin

Prepped library fragments

Add blocking oligos 1

xGen® Blocking Oligos

Blocked library fragments

Hybridize targets to capture probes 2

xGen® Lockdown® Probes

Hybridized library fragments

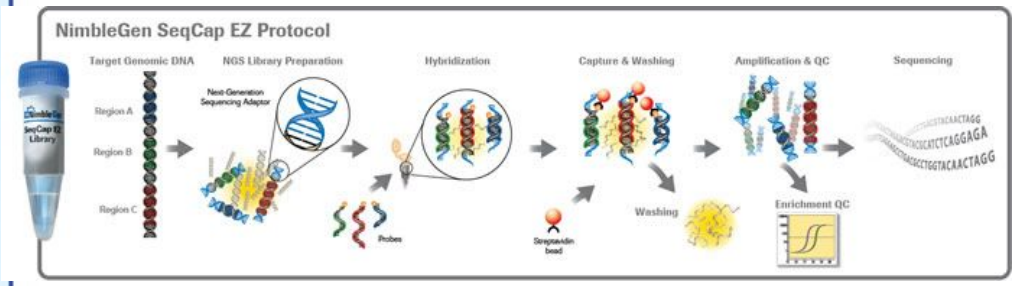
Incubate with magnetic streptavidin beads 3

Streptavidin-coated magnetic beads

Bead-bound target sequences

Isolate targets with magnet 4

IDT
INTEGRATED DNA TECHNOLOGIES



Ampliseq

Since dideoxy (Sanger) sequencing was developed over 30 years ago, amplicon sequencing has been a mainstay of genome analysis.

Now, with AmpliSeq™, it is possible to simultaneously amplify, sequence and genotype hundreds of genomic regions in a single project.

Ampliseq

By focusing next-generation DNA sequencing (NGS) technologies on specific targets, tens to hundreds of genetic markers can be quickly and cost effectively identified or genotyped in large populations.

video

Ion AmpliSeq™ Target Selection Overview

(RNA only) Reverse transcribe the RNA

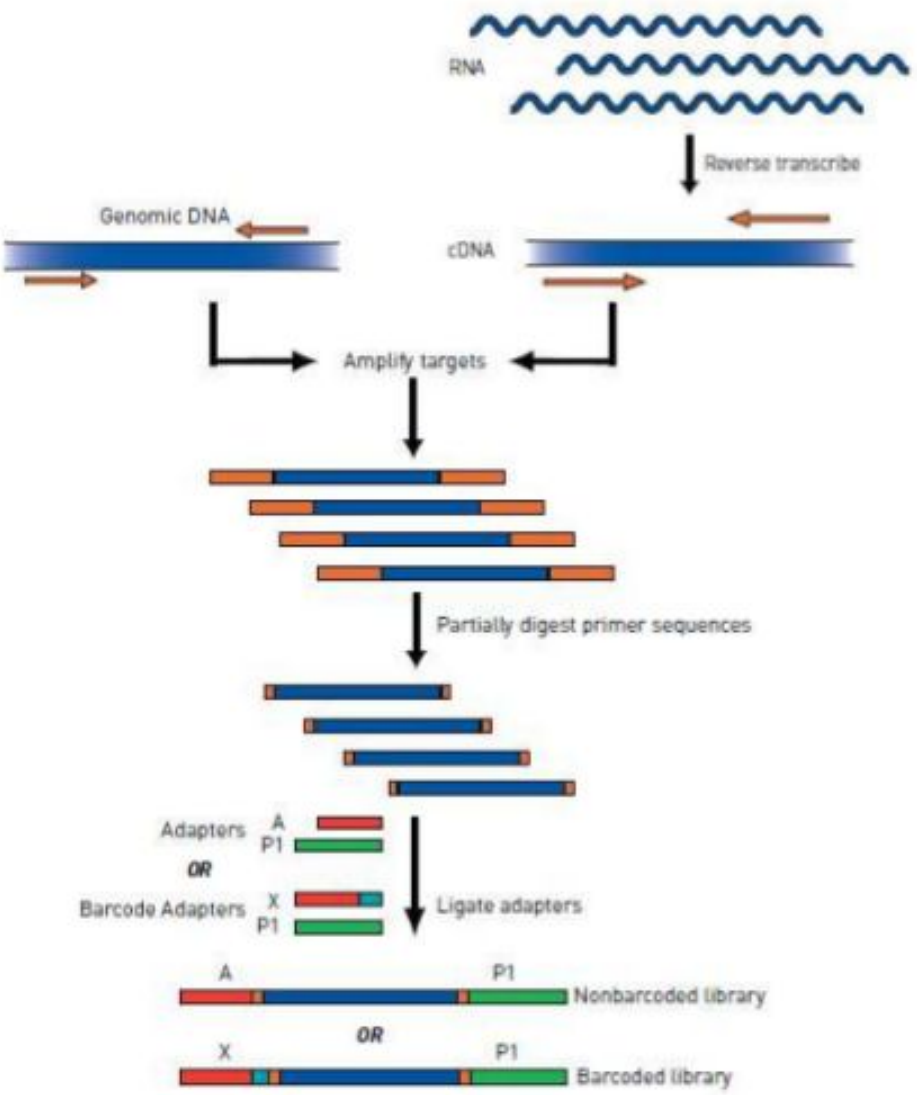
Amplify targets

Partially digest primer sequences

Ligate adapters to the amplicons and purify

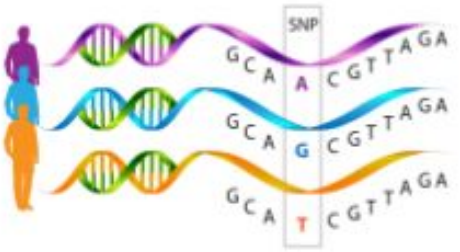
Option 1: Equalize the library (not recommended for libraries from RNA)	Option 2: Quantify the unamplified library by qPCR (recommended for libraries from RNA)	Option 3: Quantify the amplified library with the Qubit® 2.0 Fluorometer or Agilent® 2100 Bioanalyzer® instrument

Combine libraries *(optional)* and proceed to template preparation.



Key technology features:

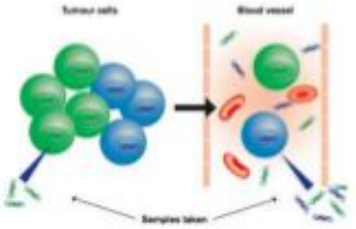
- RNA pre-calculated designs
- DNA gene designs
  - "Standard" AmpliSeq designs
- DNA hotspot designs
  - 1-pool designs for SNP's



- Designs for "Any Genome"
  - AgBio applications and more

- Design customization
  - Sub-setting
  - Mix and match
  - Whitelist

- Support for Cell Free DNA (cfDNA) designs
  - 140bp amplicon sizes now available



Ion AmpliSeq™ Panels

For DNA analysis and RNA expression measurement

Ready-to-use **Human Identity Panel**	Ready-to-use **Human Ancestry Panel**	Ready-to-use **RNA Apoptosis Panel**	Ready-to-use **RNA Cancer Panel**
Ready-to-use **Cancer Hotspot Panel**	Ready-to-use **Comprehensive Cancer Panel**	Ready-to-use **Inherited Disease Panel**	Community **RNA Fusion Lung Cancer Research Panel**
Community ***BRCA* 1 & 2 Research Panel**	Community **Hearing Loss Research Panel**	Community ***CFTR* Research Panel**	Community ***TP53* Research Panel**
Community **AML Research Panel**	Community **Cardio Research Panel***	Community **Dementia Research Panel**	Community **Colon & Lung Cancer Research Panel**

Ion AmpliSeq™ Cancer Hotspot Panel v2

As little as 10ng input
50 genes
207 amplicons

The Ion AmpliSeq™ Cancer Panel targets 50 genes

ABL1	*EZH2*	*JAK3*	*PTEN*
AKT1	*FBXW7*	*IDH2*	*PTPN11*
ALK	*FGFR1*	*KDR*	*RB1*
APC	*FGFR2*	*KIT*	*RET*
ATM	*FGFR3*	*KRAS*	*SMAD4*
BRAF	*FLT3*	*MET*	*SMARCB1*
CDH1	*GNA11*	*MLH1*	*SMO*
CDKN2A	*GNAS*	*MPL*	*SRC*
CSF1R	*GNAQ*	*NOTCH1*	*STK11*
CTNNB1	*HNF1A*	*NPM1*	*TP53*
EGFR	*HRAS*	*NRAS*	*VHL*
ERBB2	*IDH1*	*PDGFRA*	
ERBB4	*JAK2*	*PIK3CA*	

Ion AmpliSeq™ Comprehensive Cancer Panel (CCP)

- Targets coding exons in 409 human oncogenes and tumor suppressor genes
- ~16,000 amplicons
- Detection of known COSMIC somatic mutations

Ion AmpliSeq™ Exome

- 293,903 primer pairs across 12 primer pools
  - >24,500-plex PCR!
  - ~2.4 M PCRs per plate of 8 exomes!!
- Total DNA input as low as 50ng
- Covers >97% of CCDS (Release 12)
  - >19,000 coding genes >198,000 coding exons (no UTRs, miRNAs, or ncRNAs)
  - ~85% of human disease-causing variants found in coding regions or splice junctions

- Amplicon size range 225-275 bp
  - Average insert size is ~202 bp

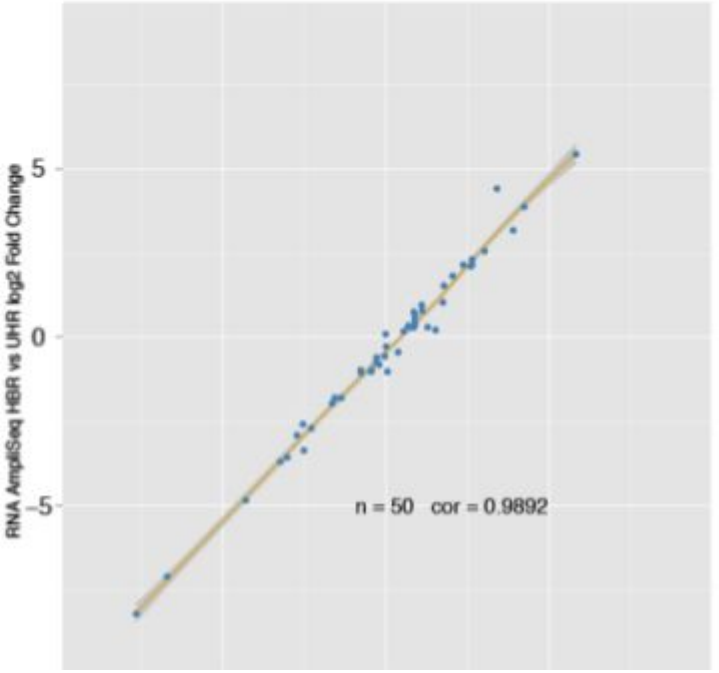
Ion AmpliSeq™ RNA Cancer Panel

50 genes, from 500 pg unfixed RNA, one tube

ABL1	EZH2	JAK3	PTEN	AKT1
FBXW7	IDH2	PTPN11	ALK	FGFR1
KDR	RB1	APC	FGFR2	KIT
RET	ATM	FGFR3	KRAS	SMAD4
BRAF	FLT3	MET	SMARCB1	CDH1
GNA11	MLH1	SMO	CDKN2A	GNAS
MPL	SRC	CSF1R	GNAQ	NOTCH1
STK11	CTNNB1	HNF1A	NPM1	TP53
EGFR	HRAS	NRAS	VHL	ERBB2
IDH1	PDGFRA	ERBB4	JAK2	PIK3CA

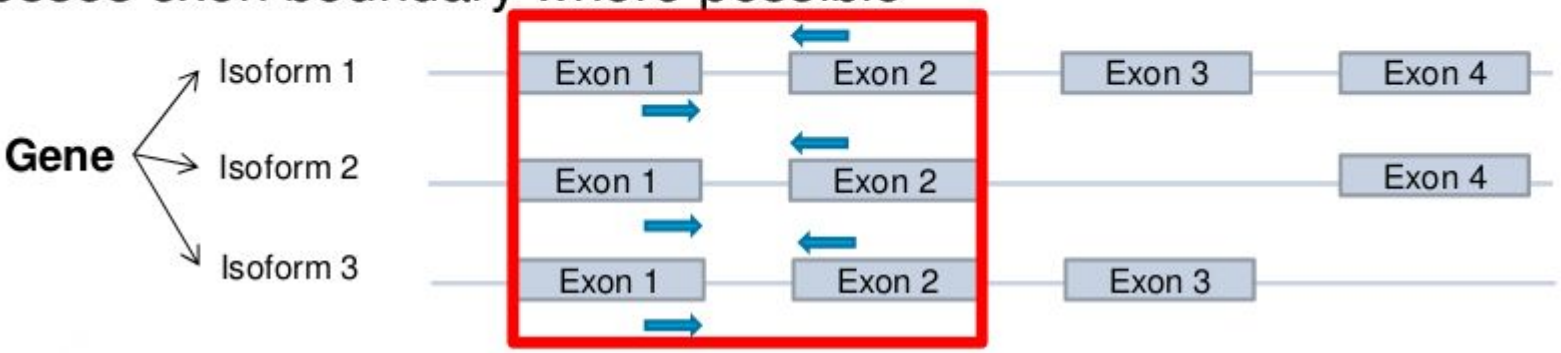
Targeted quantitative expression

Complement to Cancer Hotspot Panel



Ion AmpliSeq™ Transcriptome – Overview

- **Comprehensive** coverage of RefSeq - 20,802 genes targeted
  - Single primer pool – simple workflow
  - Reports at **gene-level**
- **FFPE-compatible** as little as 10 ng RNA
  - Total RNA input from FFPE and other sources – **no selection or enrichment**
- Assay design
  - One amplicon per gene
    - ~150 bp amplicon size with ~110 bp insert size
  - Crosses exon boundary where possible



Benefits of Target Enrichment vs. Amplicon Sequencing

Target Enrichment	Amplicon Sequencing
Larger gene content, typically > 50 genes	Smaller gene content, typically < 50 genes
More comprehensive profiling for all variant types	Ideal for analyzing single nucleotide variants and insertions/deletions (indels)
More comprehensive method, but with longer hands-on time and turnaround time	More affordable, easier workflow


DNA Rearrangements and Markers
RAD
PE RAD-Seq
dsRAD-seq
HydEn-seq
Ribose-seq
G4-seq
CAP-seq
CPT-Seq
RC-Seq
Tiv-Seq
3hSeq
TC-Seq
3g-seq
Rap-Seq
EC-seq
Bubble-Seq
DNA Low-Level Detection
smMIP
MIPSTR
MDA
3MS-HDA
MALBAC
Nuc-seq
SNSS
OS-Seq
Duplex-Seq
GAT-seq
Key
Epigenetics
BS-Seq
BisulFlex-seq
WGBS
EpiGnome™
HELP-Seq
ecBS-seq
PBAT
BSPP
RRBS
ecRRBS
BSAS
Methyl-Seq
MRE-Seq
T-WGBS
3BP1-seq
Aba-seq
TAmC-Seq
fC-Seal
CAB-Seq
oxBS-Seq
RedBS-Seq
fCAB-Seq
HAB-seq
RRMAB-seq
TAB-Seq
MIRA
MeDIP-Seq
DIP-seq
hMeDIP-Seq
MBDCap-seq
Methyl-Cap-seq
MBD-Seq
MiGS
BisChIP-Seq
ChIP-BS-seq
DNA-Protein Interactions
DamID
DNase-Seq
DNaseI-Seq
MAINE-Seq
MNase-Seq
Nucleo-Seq
X-ChIP
ORGANIC
CATCH-IT
ChIP-Seq
ChIP-exo
HT-ChIP
Histone methylation
Chem-seq
FAIRE-seq
Sono-Seq
NOMe-Seq
ATAC-Seq
Chia-PET
Hi-C
3-C
Capture-C
4-C
5-C
PB-seq
Pu-seq
Break-seq
SELEX
SELEX-seq
HT-SELEX
HITS-FLIP
Protein-Protein Interactions
PD-Seq
PnpP-PD
PDZ-Seq

LOW level DNA detection

Single-cell genomics can be used to identify and study circulating tumor cells, cell-free DNA, microbes, uncultured microbes, for preimplantation diagnosis, and to help us better understand tissue-specific cellular differentiation.

DNA replication during cell division is not perfect; as a result, progressive generations of cells accumulate unique somatic mutations.

Consequently, each cell in our body has a unique genomic signature, which allows the reconstruction of cell lineage trees with very high precision.

These cell lineage trees can predict the existence of small populations of stem cells. This information is important for fields as diverse as cancer development preimplantation, and genetic diagnosis.

Reviews

Blainey P. C. (2013) The future is now: single-cell genomics of bacteria and archaea. FEMS Microbiol Rev 37: 407-427

Lovett M. (2013) The applications of single-cell genomics. Hum Mol Genet 22: R22-26

Shapiro E., Biezuner T. and Linnarsson S. (2013) Single-cell sequencing-based technologies will revolutionize whole-organism science. Nat Rev Genet 14: 618-630

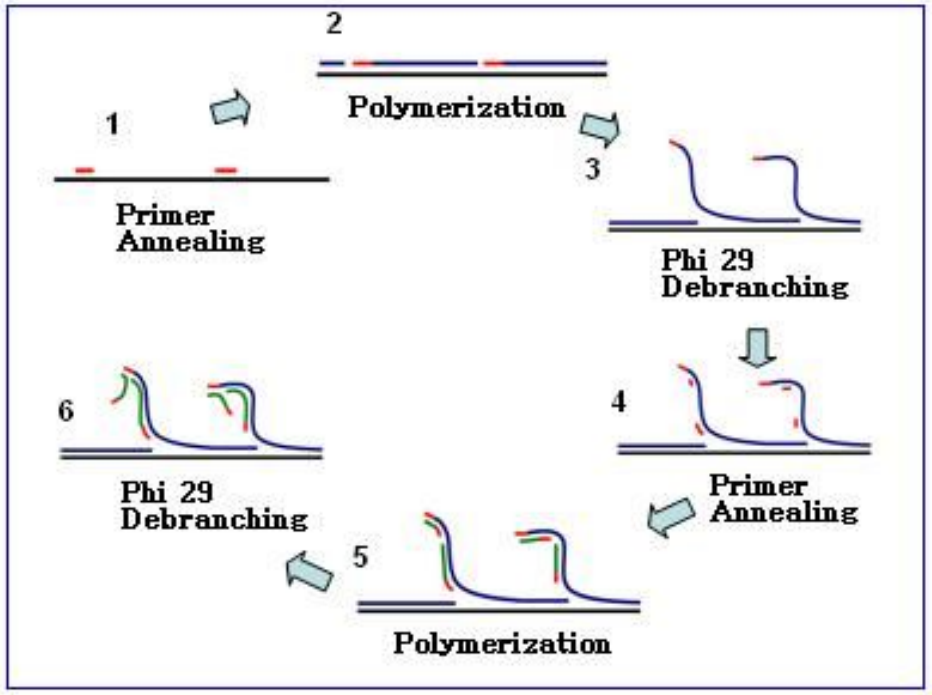
Low-Level DNA Detection

- MULTIPLE DISPLACEMENT AMPLIFICATION (MDA)

Multiple displacement amplification (MDA) is a method commonly used for sequencing microbial genomes due to its ability to amplify templates larger than 0.5 Mbp, but it can also be used to study genomes of other sizes.

- In this method, 3'-blocked random hexamer primers are hybridized to the template, followed by synthesis with Phi 29 polymerase.
- Phi 29 performs strand-displacement DNA synthesis, allowing for efficient and rapid DNA amplification.

WGA_MDA



MDA can generate 1–2 μg of DNA from single cell with genome coverage of up to 99%.

General work flow of MDA:

- Sample preparation: Samples are collected and diluted in the appropriate reaction buffer (Ca2+ and Mg2+ free).
- Condition: The MDA reaction with Φ29 polymerase is carried out at 30 °C. The reaction usually takes about 2.5–3 hours.
- End of reaction: Inactivate enzymes at 65 °C before collection of the amplified DNA products
- DNA products can be purified with commercial purification kit.

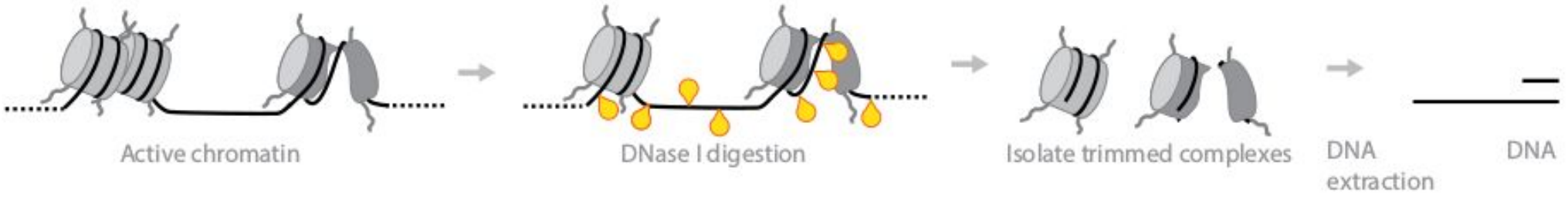
DNA-PROTEIN INTERACTIONS

Chromatin remodeling is a dynamic process driven by factors that change DNA-protein interactions. These epigenetic factors can involve protein modifications, such as histone methylation, acetylation, phosphorylation, and ubiquitination.

Histone modifications determine gene activation by recruiting regulatory factors and maintaining an open or closed chromatin state. Epigenetic factors play roles in tissue development, embryogenesis, cell fate, immune response,

DNASE I HYPERSENSITIVE SITES SEQUENCING (DNASE-SEQ)

Sequences bound by regulatory proteins are protected from DNase I digestion. Deep sequencing provides accurate representation of the location of regulatory proteins in genome.

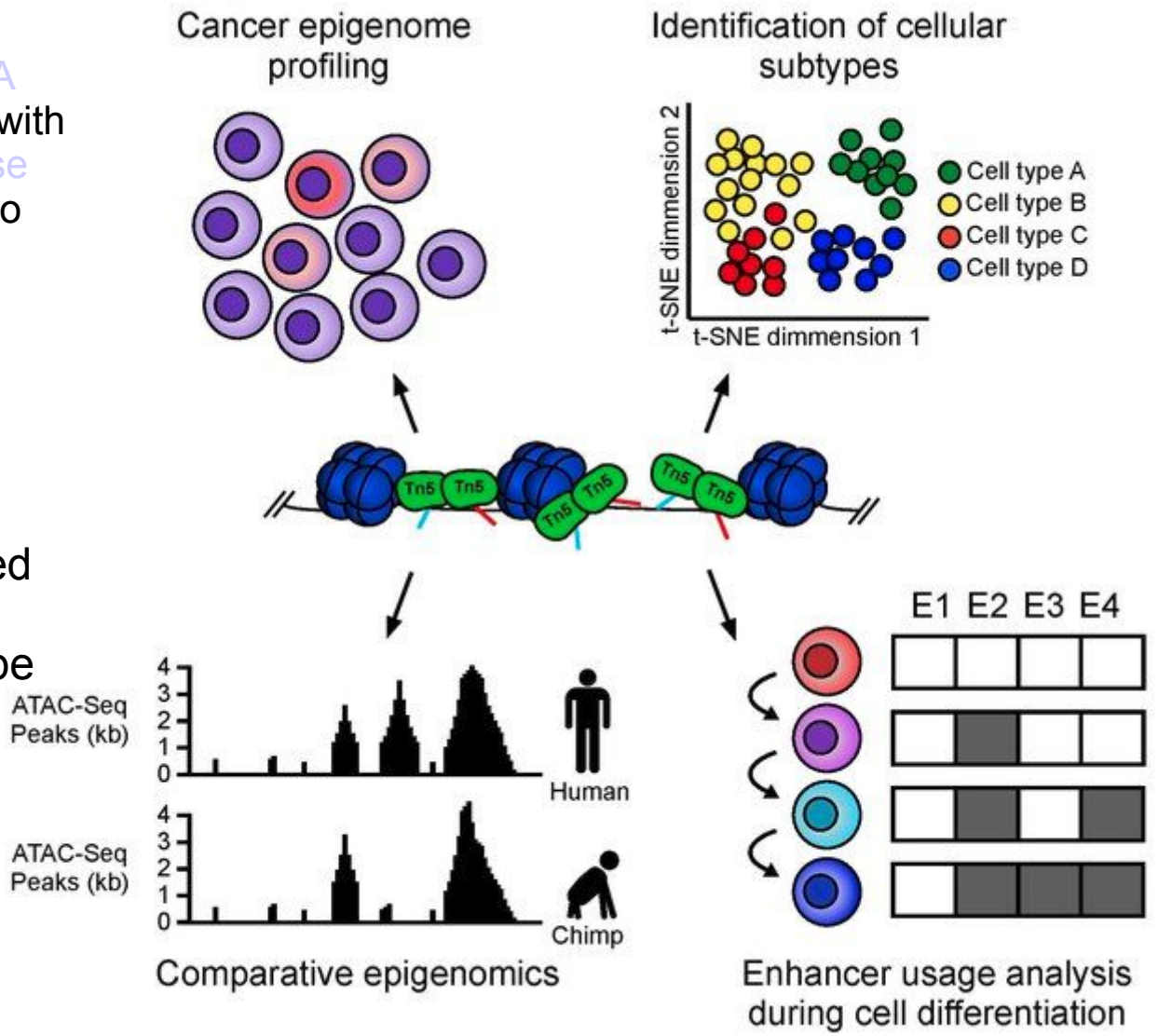


ATAC-seq

Assay for Transposase-Accessible Chromatin using sequencing

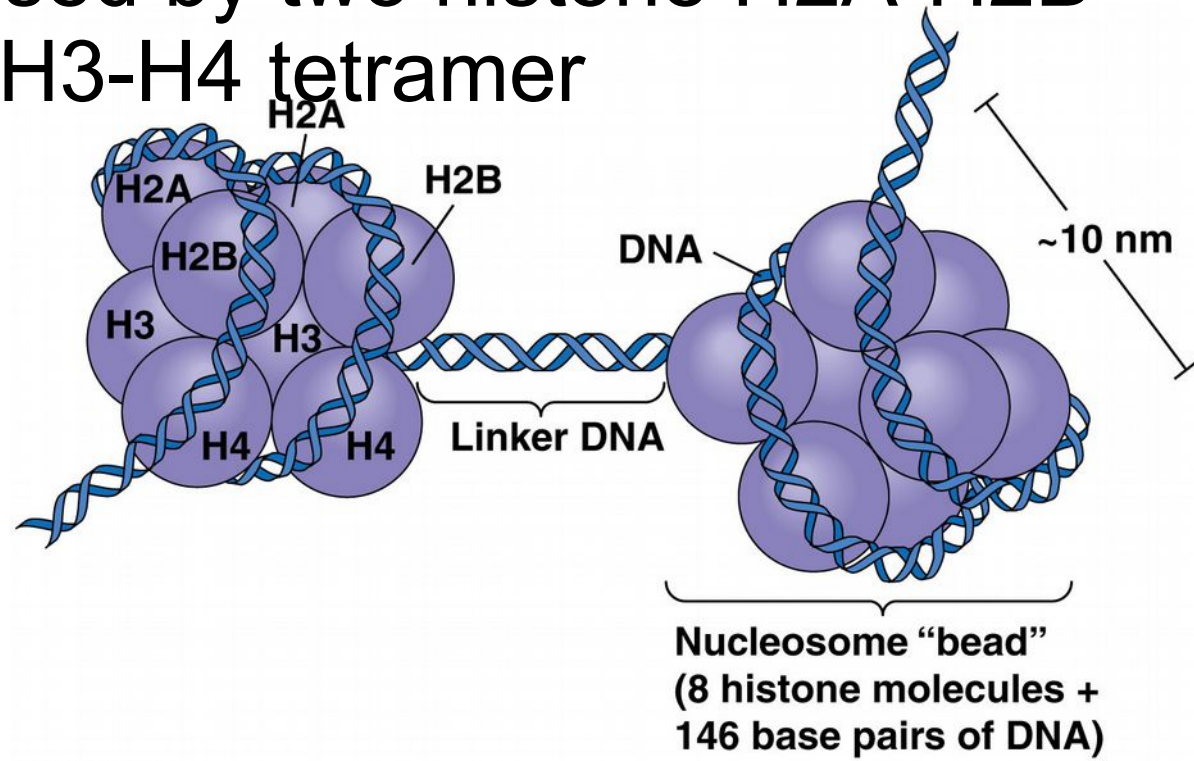
ATAC-seq identifies accessible DNA regions by probing open chromatin with hyperactive mutant Tn5 Transposase that inserts sequencing adapters into open regions of the genome.

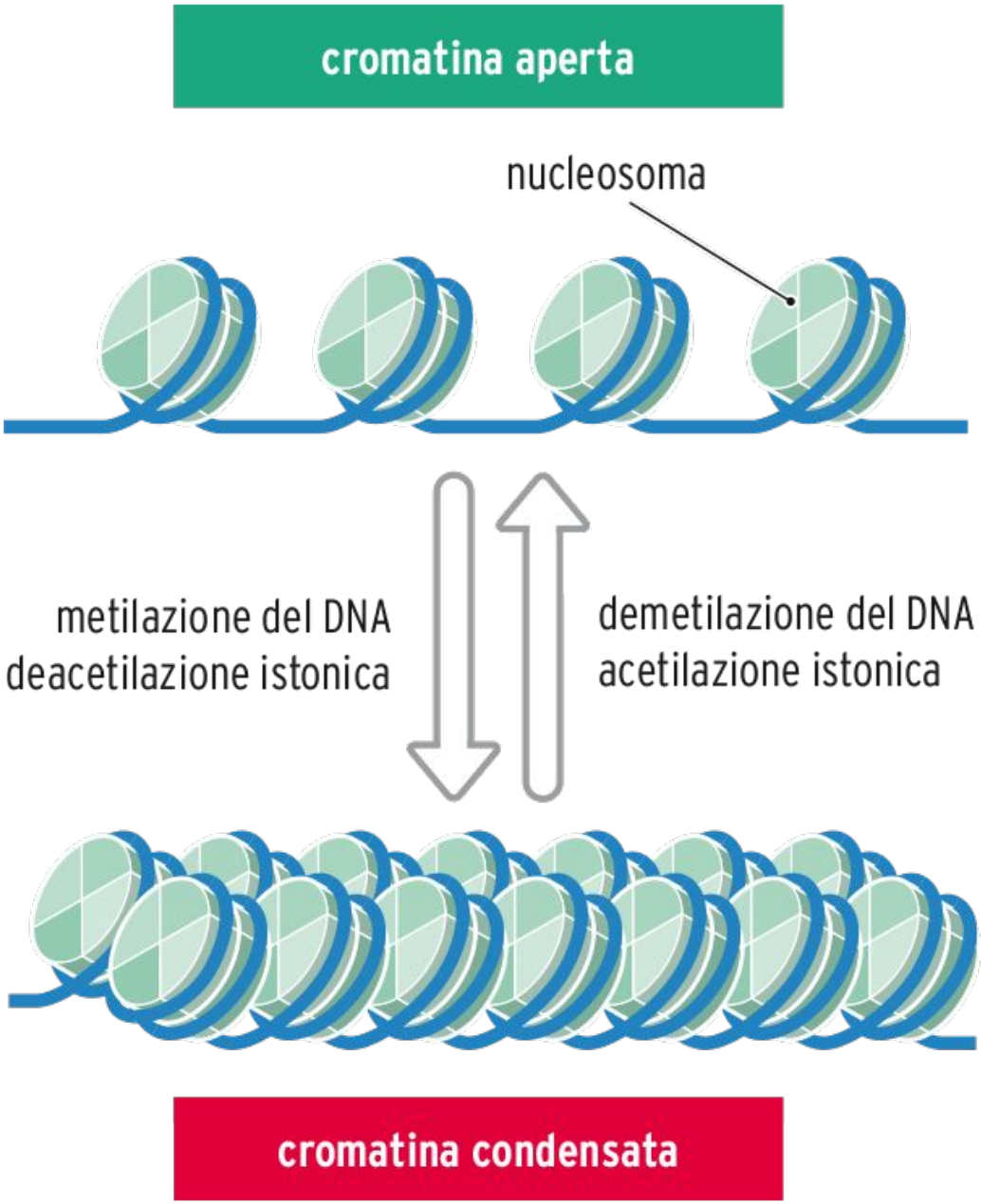
ATAC-Seq has also been applied to defining the genome-wide chromatin accessibility landscape in human cancers,



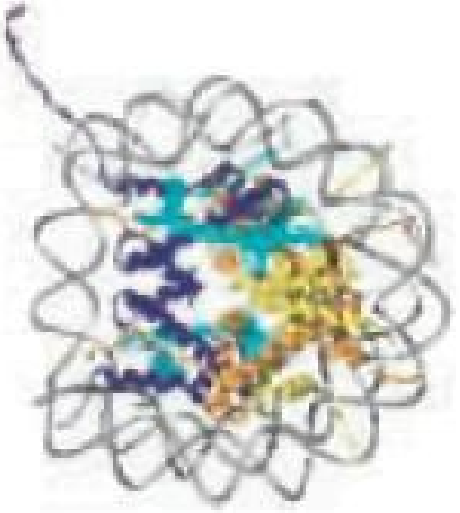
Chip-seq

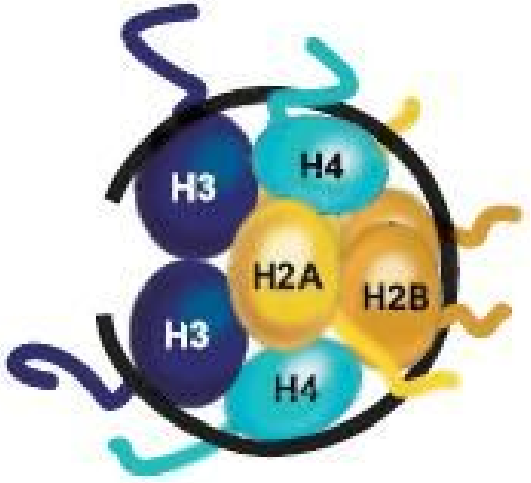
In eukaryotic cells, the genome is highly organized within the nucleus in a complex compact structure known as chromatin. The basic unit of chromatin is the nucleosome, which consists of 146 bp of DNA wrapped around the four histone proteins arranged as an octamer composed by two histone H2A-H2B dimers and a histone H3-H4 tetramer




cromatina aperta
nucleosoma
metilazione del DNA
deacetilazione istonica
demetilazione del DNA
acetilazione istonica
cromatina condensata

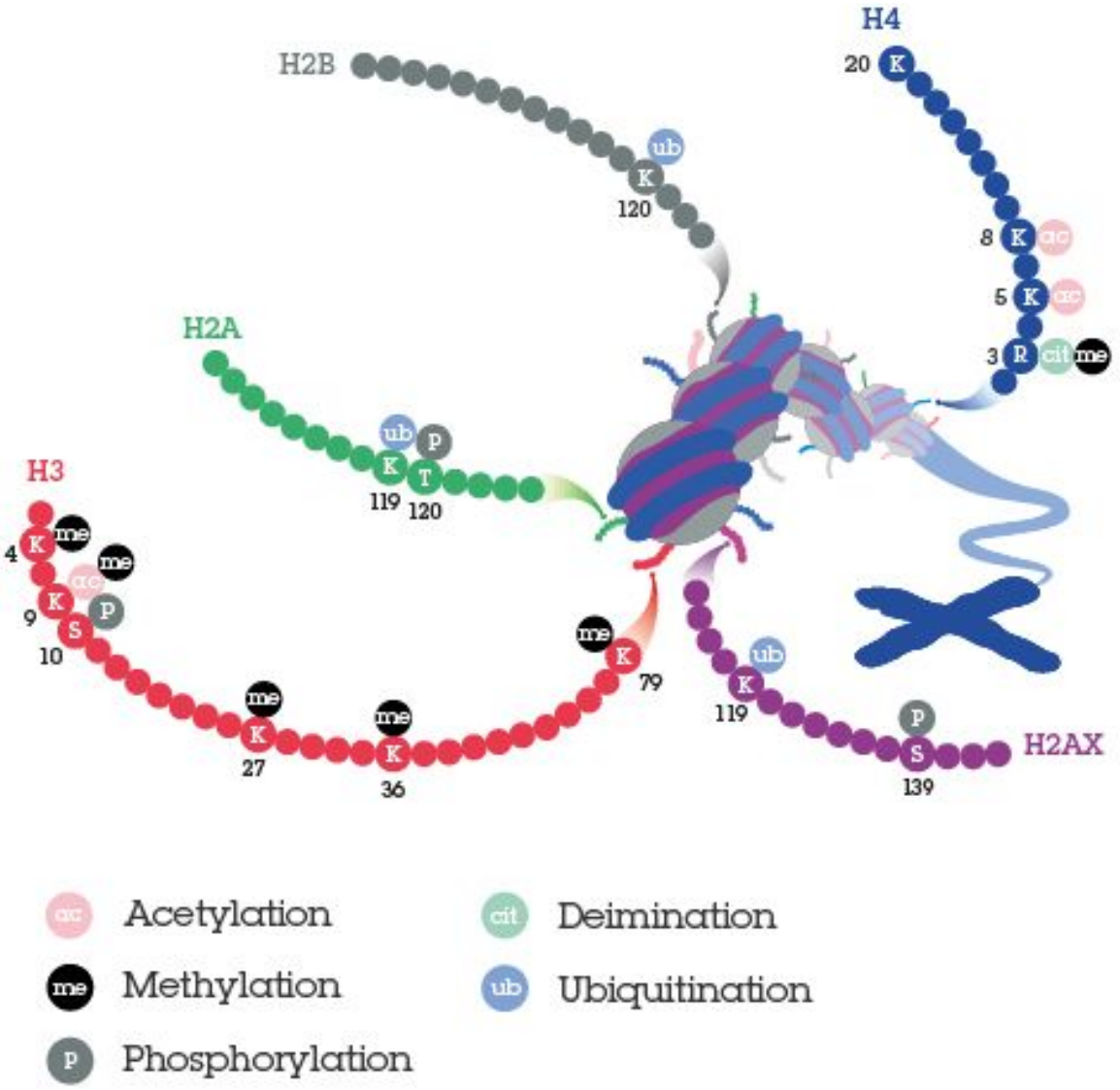
Each histone protein contain the so called histone fold structural motif, flanked by unstructured N- and C-terminal tails, ranging from 15 (H2A) to 35 (H3) amino acids, that protrude from the nucleosomal core

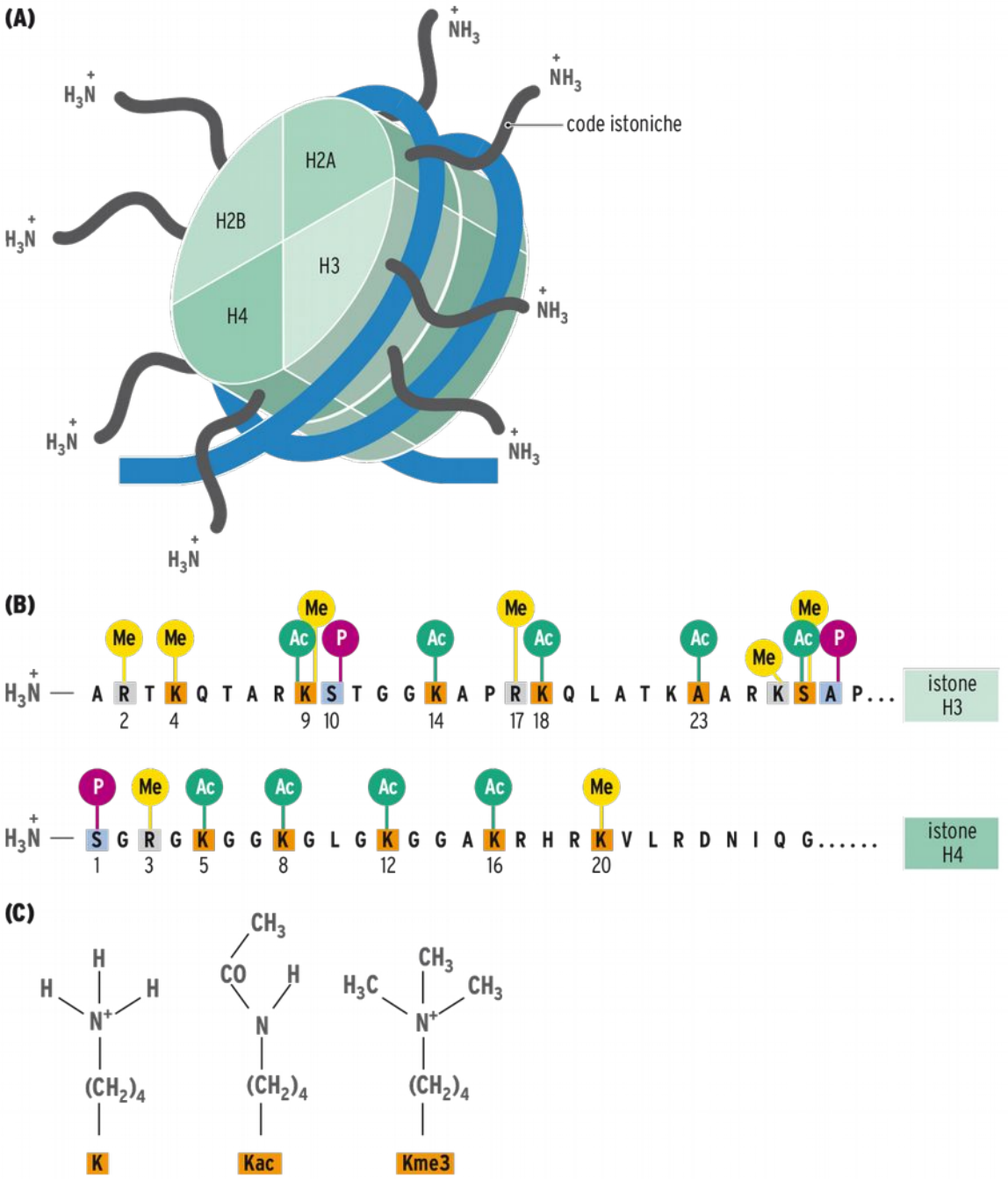




Histone modifications

Acetylation and methylation of specic lysine or arginine residues in histones H3 and H4 are reversible and have been associated to gene transcription regulation

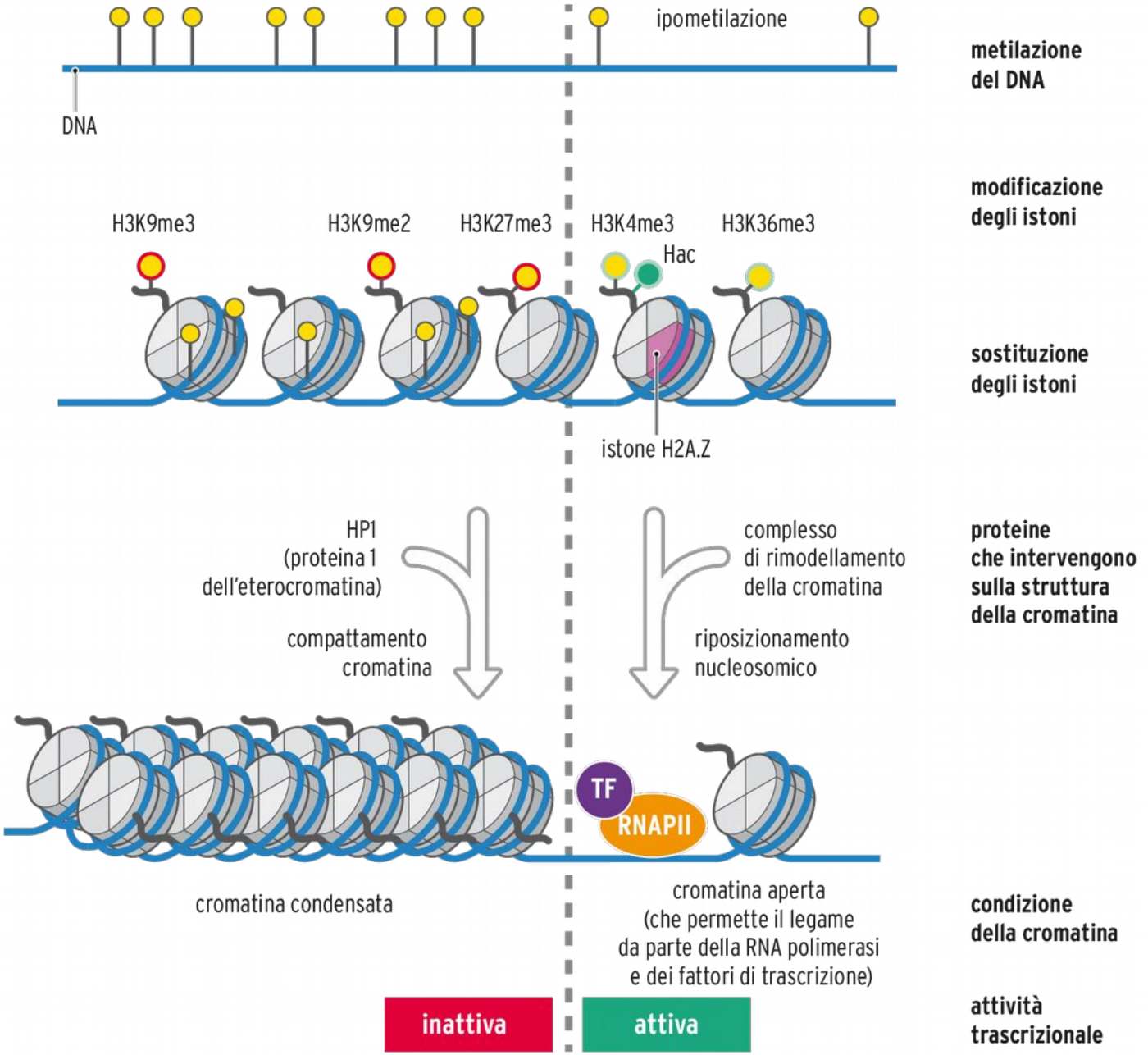



(A)
NH_3^+
NH_3^+
H_3N^+
code istoniche
H2A
H2B
H3
H4
H_3N^+
NH_3^+
H_3N^+
NH_3^+
H_3N^+
(B)
H_3N^+ — A R T K Q T A R K S T G G K A P R K Q L A T K A A R K S A P...
2 4 9 10 14 17 18 23
Me Me Ac P Ac Me Ac Ac Me Ac P
istone H3
H_3N^+ — S G R G K G G K G L G K G G A K R H R K V L R D N I Q G.....
1 3 5 8 12 16 20
P Me Ac Ac Ac Ac Me
istone H4
(C)
K
Kac
Kme3

H3K4, H3K36, and H3K79 methylations are generally linked to active gene expression

H3K9, H3K27, and H4K20 di- and tri-methylations have been associated with gene silencing.

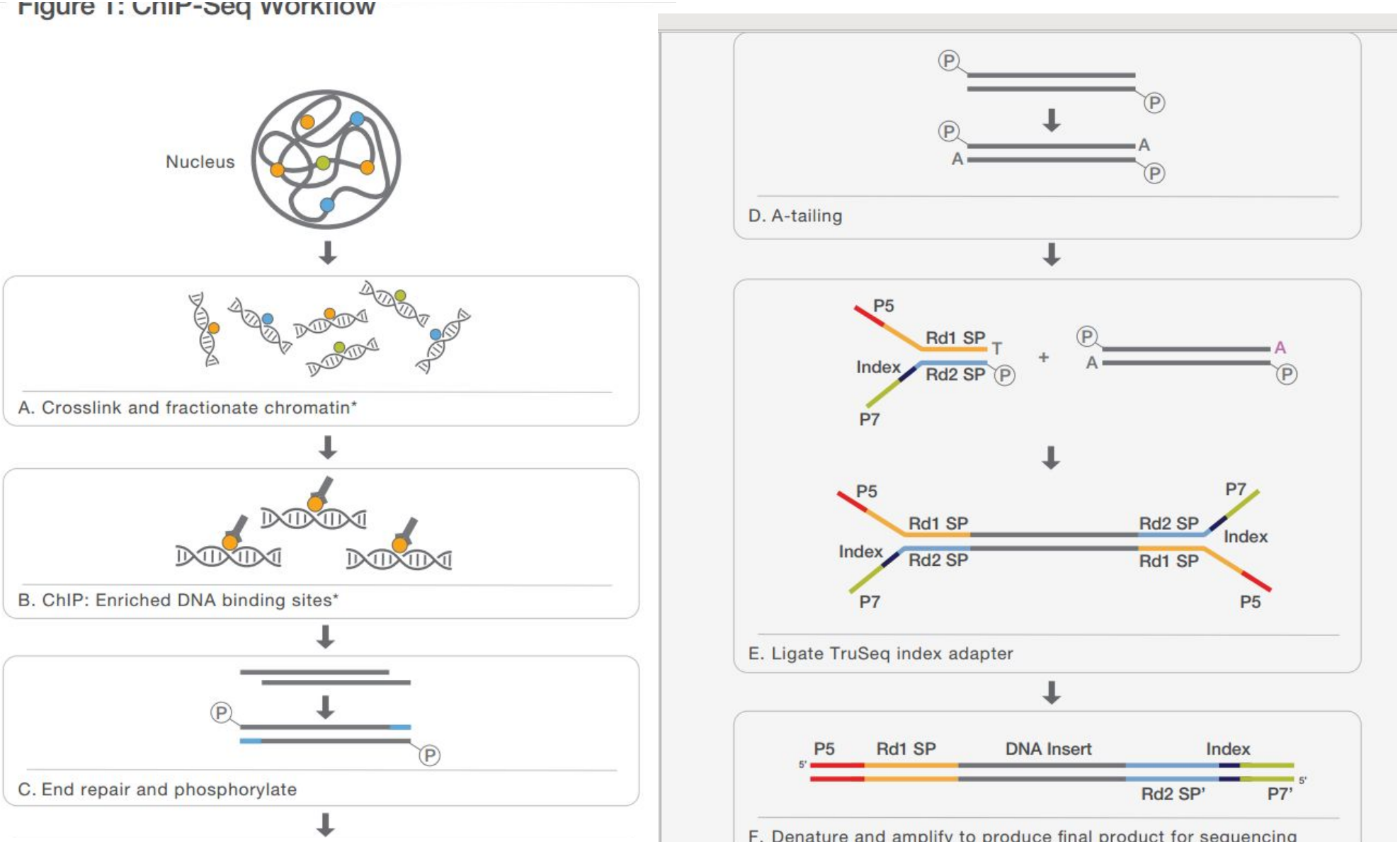
H3K4me2 and H3K4me3 have been both found predominantly on active loci, although H3K4me3 is associated with active genes, H3K4me2 can be present also in inactive genes


ipometilazione
DNA
metilazione del DNA
H3K9me3
H3K9me2
H3K27me3
H3K4me3
H3K36me3
Hac
modificazione degli istoni
sostituzione degli istoni
istone H2A.Z
HP1 (proteina 1 dell'eterocromatina)
compattamento cromatina
complesso di rimodellamento della cromatina
riposizionamento nucleosomico
proteine che intervengono sulla struttura della cromatina
TF
RNAPII
cromatina condensata
cromatina aperta (che permette il legame da parte della RNA polimerasi e dei fattori di trascrizione)
condizione della cromatina
inattiva
attiva
attività trascrizionale

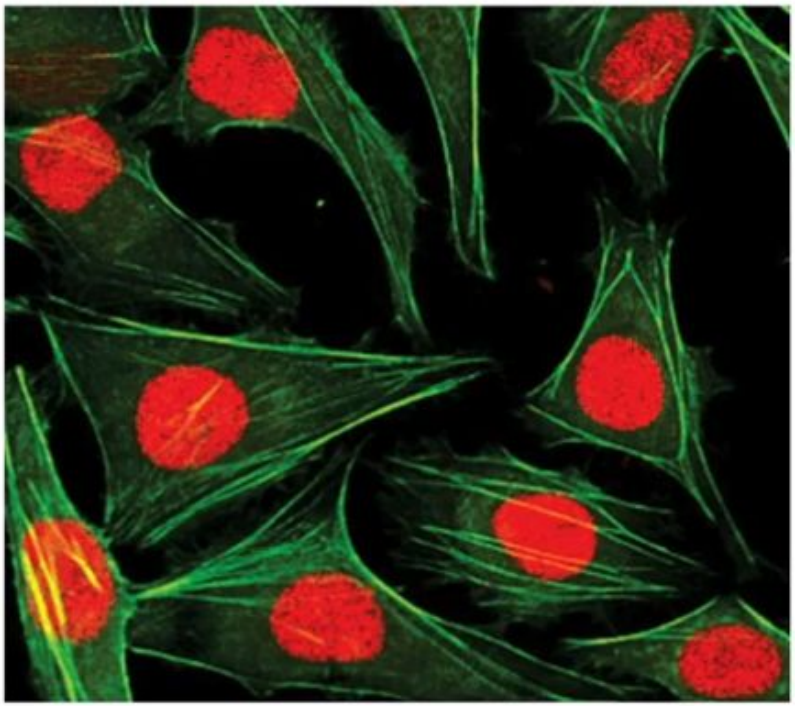
TruSeq® ChIP Sample Preparation

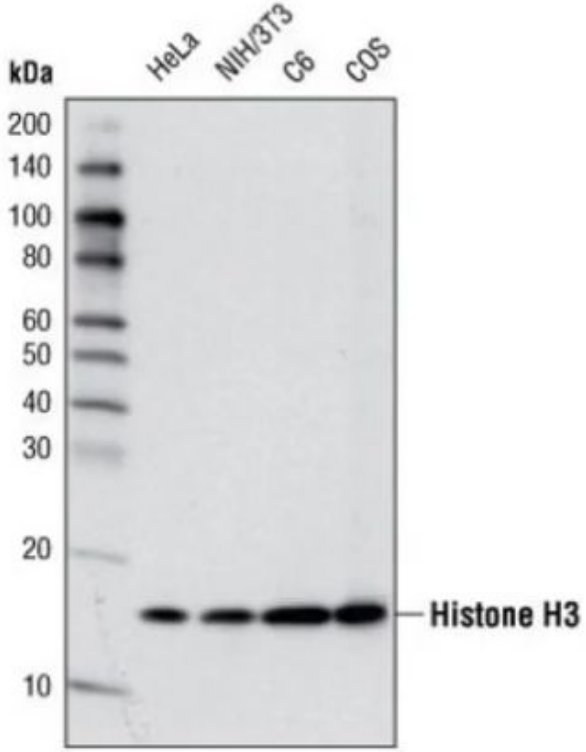
Determining how protein–DNA interactions regulate gene expression is essential for fully understanding many biological processes and disease states. This epigenetic information is complementary to DNA sequencing, genotyping, gene expression, and other forms of genomic analysis. Chromatin immunoprecipitation sequencing (ChIP-Seq) leverages next-generation sequencing (NGS) to quickly and efficiently determine the distribution and abundance of DNA-bound protein targets of interest across the genome. ChIP-Seq has become one of the most widely applied NGS-based applications, enabling researchers to reliably identify binding sites of a broad range of targets across the entire genome with high resolution and without constraints.

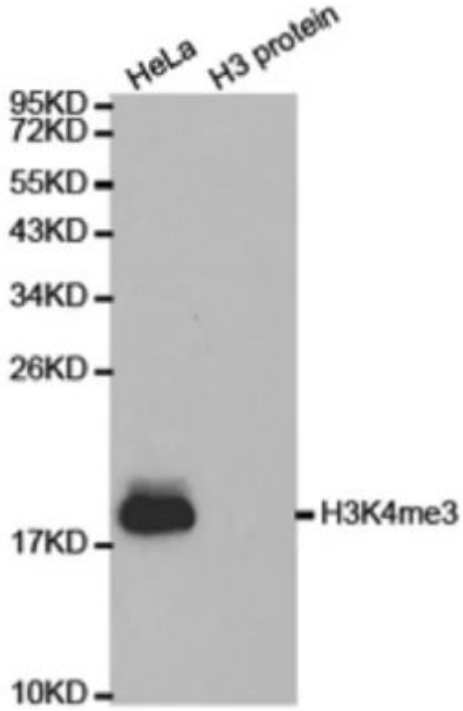
Figure 1: ChIP-Seq Workflow



Histone H3 Polyclonal Antibody

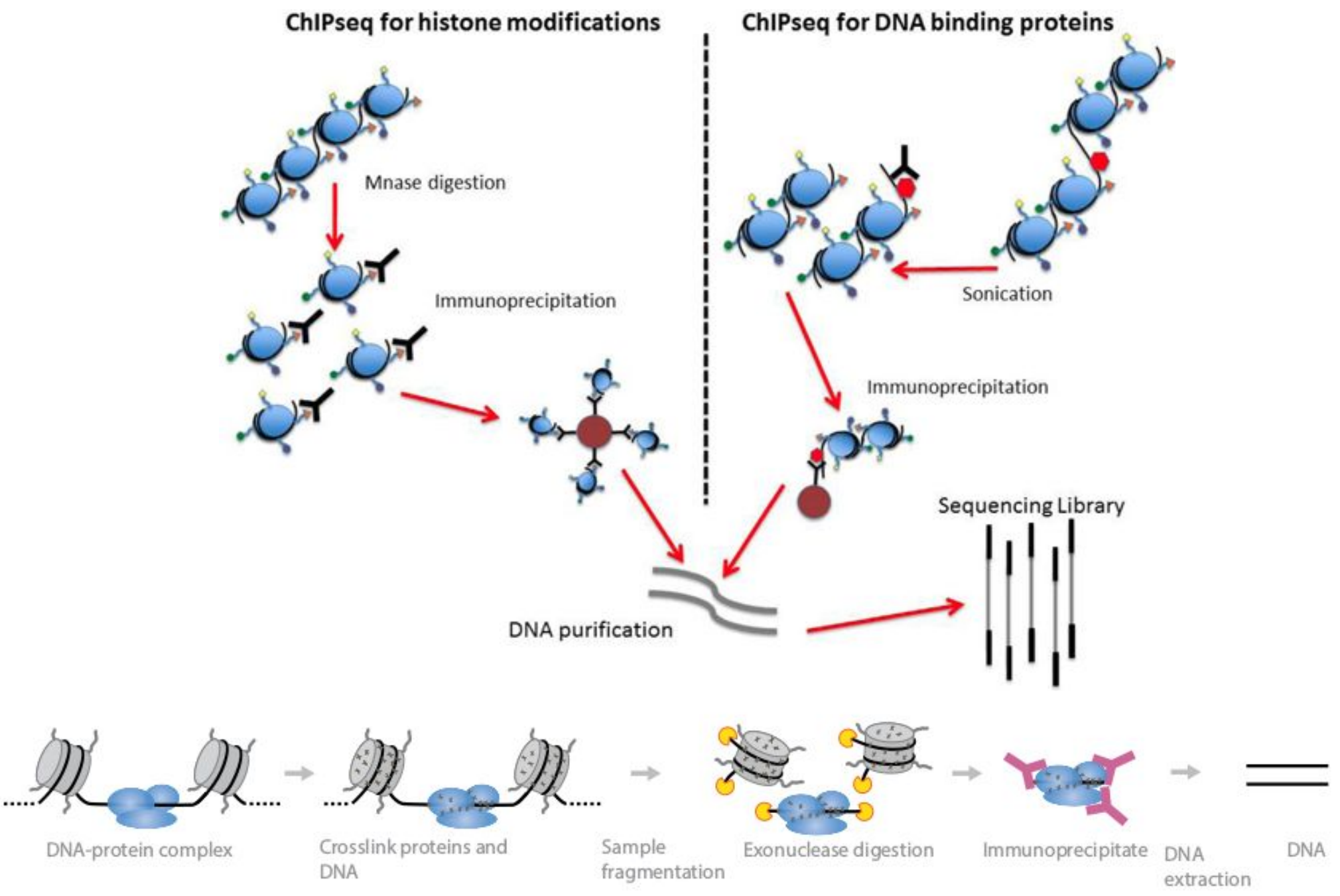




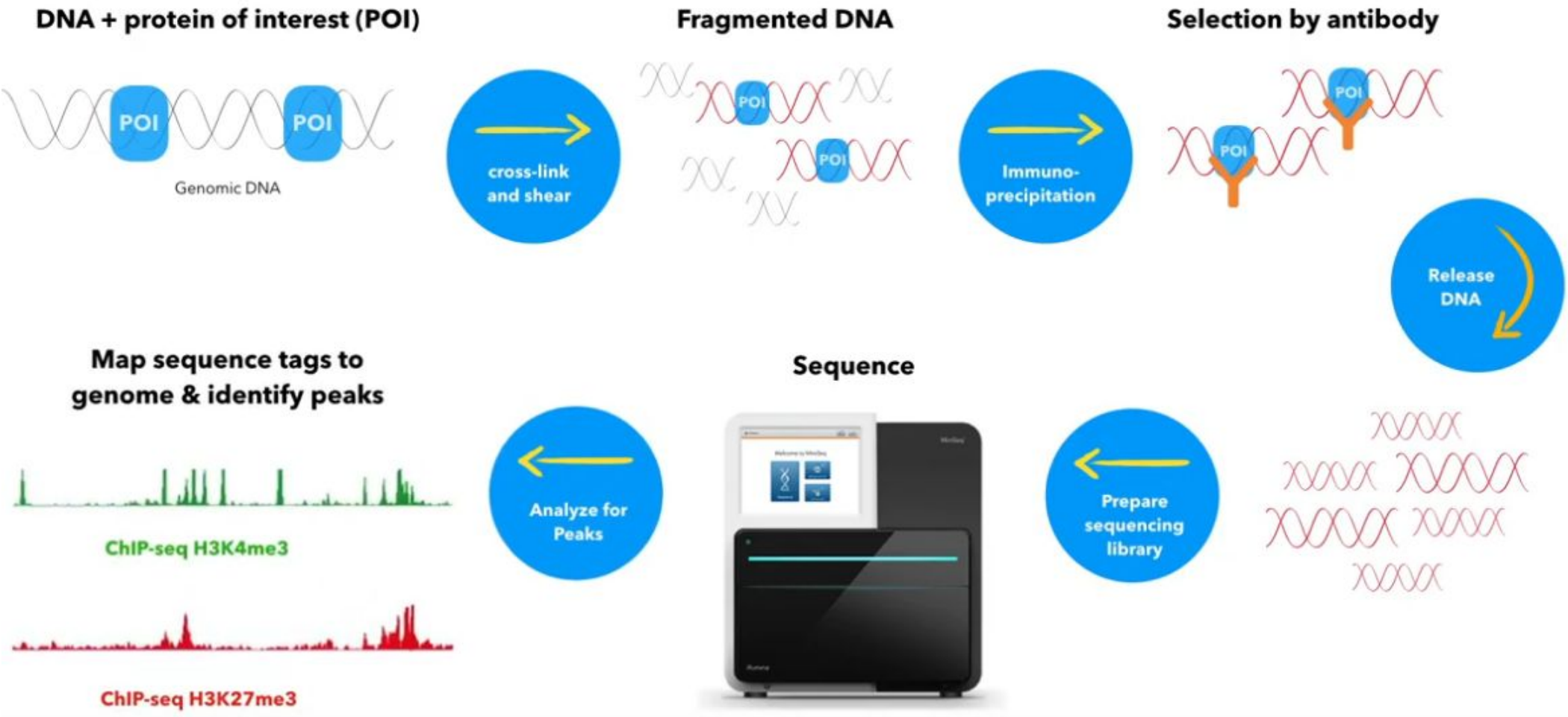


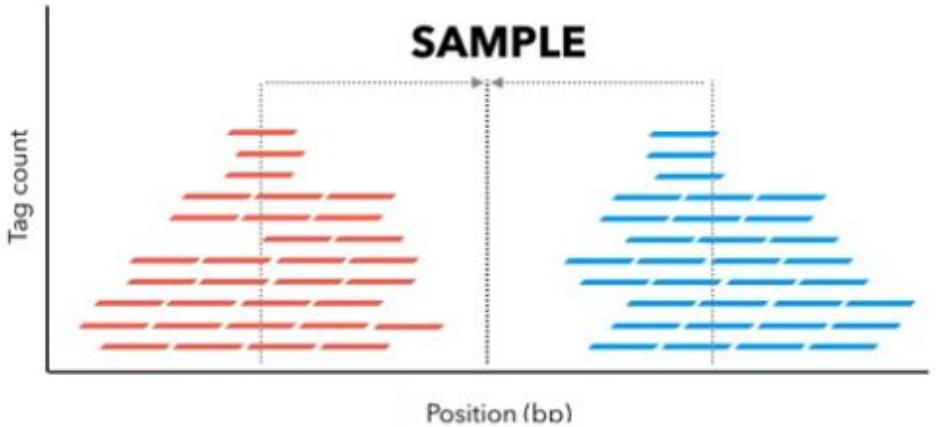
https://www.thermofisher.com/antibody/product/Histone-H3-Antibody-Polyclonal/PA5-17697

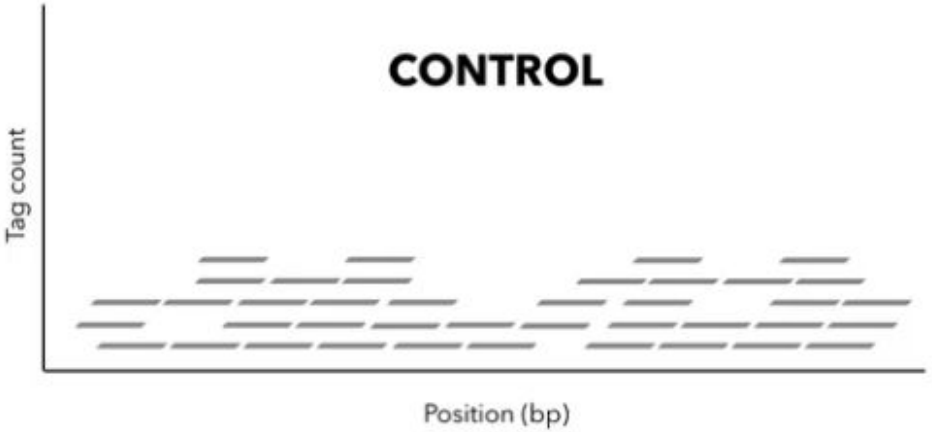
https://www.cellsignal.com/1/1/2390-monoclonal-antibody-histone-h3-d1h2-xp-rabbit-mab-uniprot-id-p68431-entrez-id-8350-4499.html

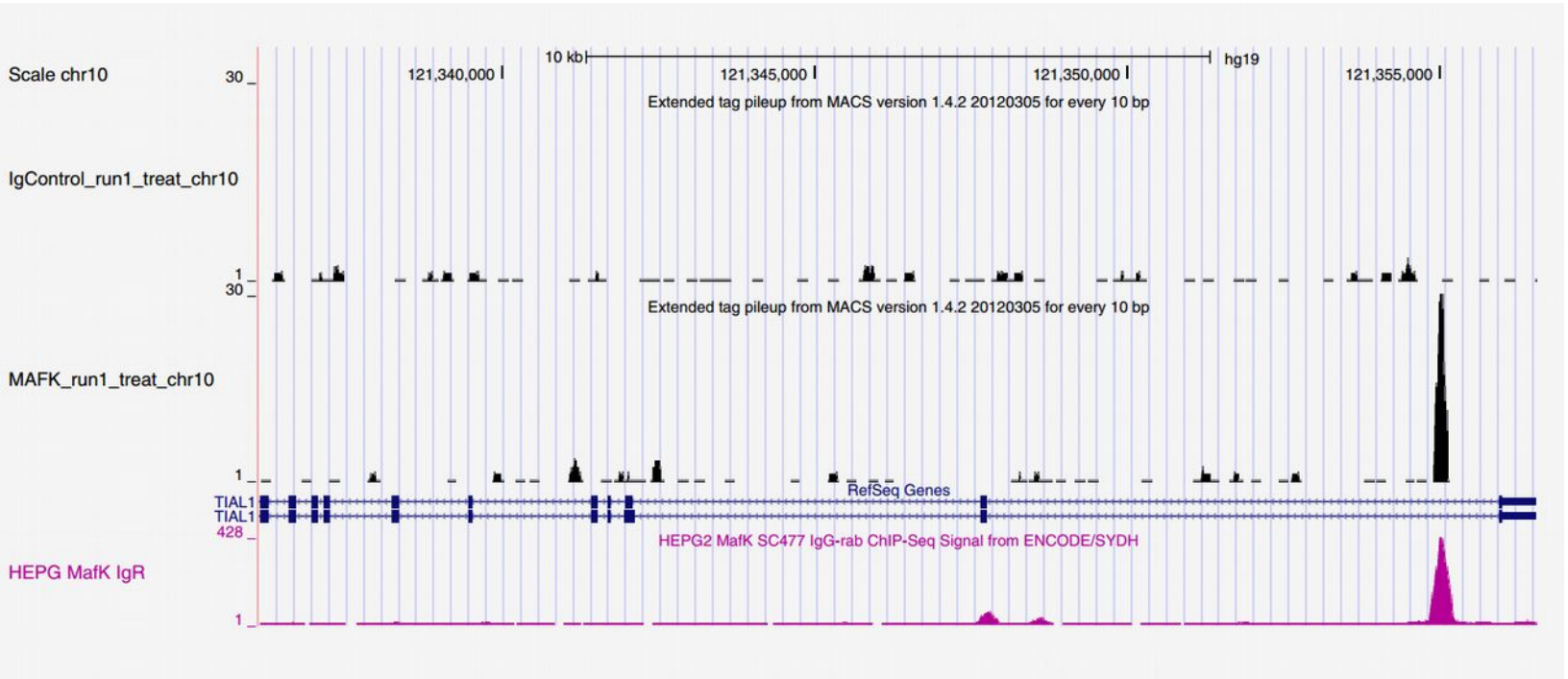


https://www.youtube.com/watch?v=4NRkoj1WFyU


DNA + protein of interest (POI)
POI
POI
Genomic DNA
cross-link and shear
Fragmented DNA
POI
POI
Immuno-precipitation
Selection by antibody
POI
POI
Release DNA
Map sequence tags to genome & identify peaks
ChIP-seq H3K4me3
ChIP-seq H3K27me3
Analyze for Peaks
Sequence
Prepare sequencing library

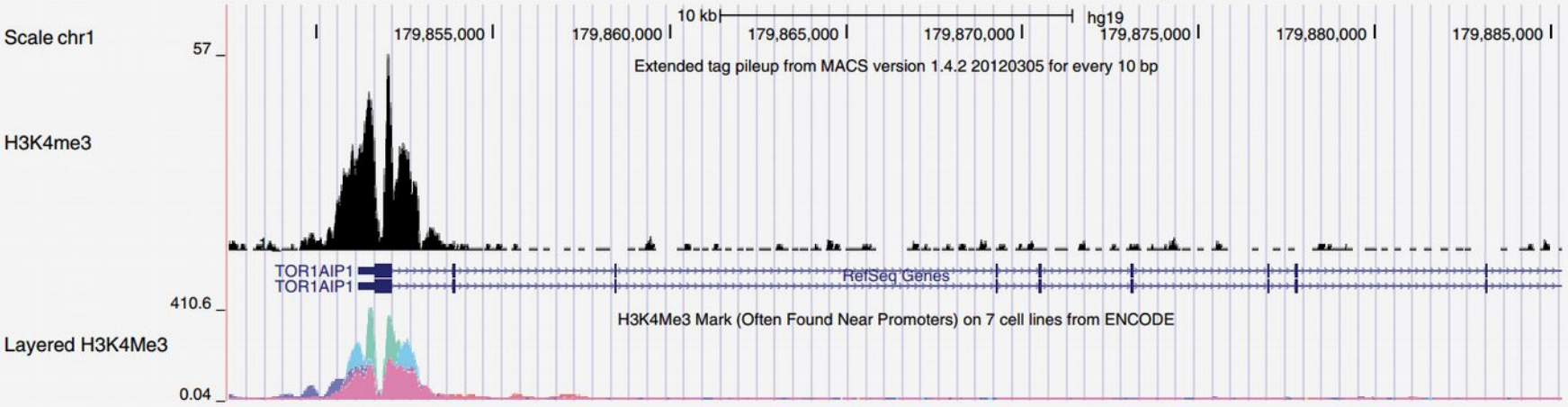

SAMPLE
Tag count
Position (bp)


CONTROL
Tag count
Position (bp)



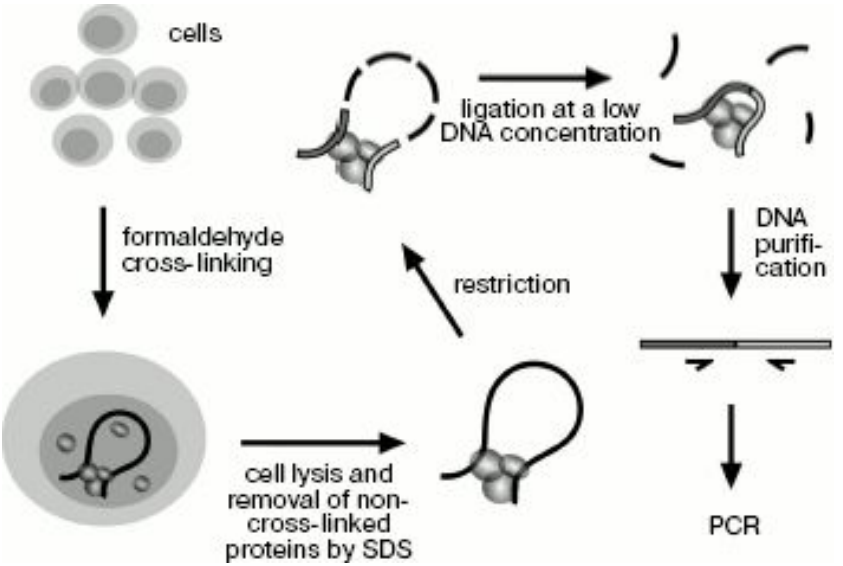
TruSeq ChIP Sample Preparation Kits enable the generation of libraries across a broad range of study designs. Above is peak data for a negative Ig control, the transcription factor target MafK, and a reference peak for MafK from the ENCODE database.

https://en.wikipedia.org/wiki/MAFK

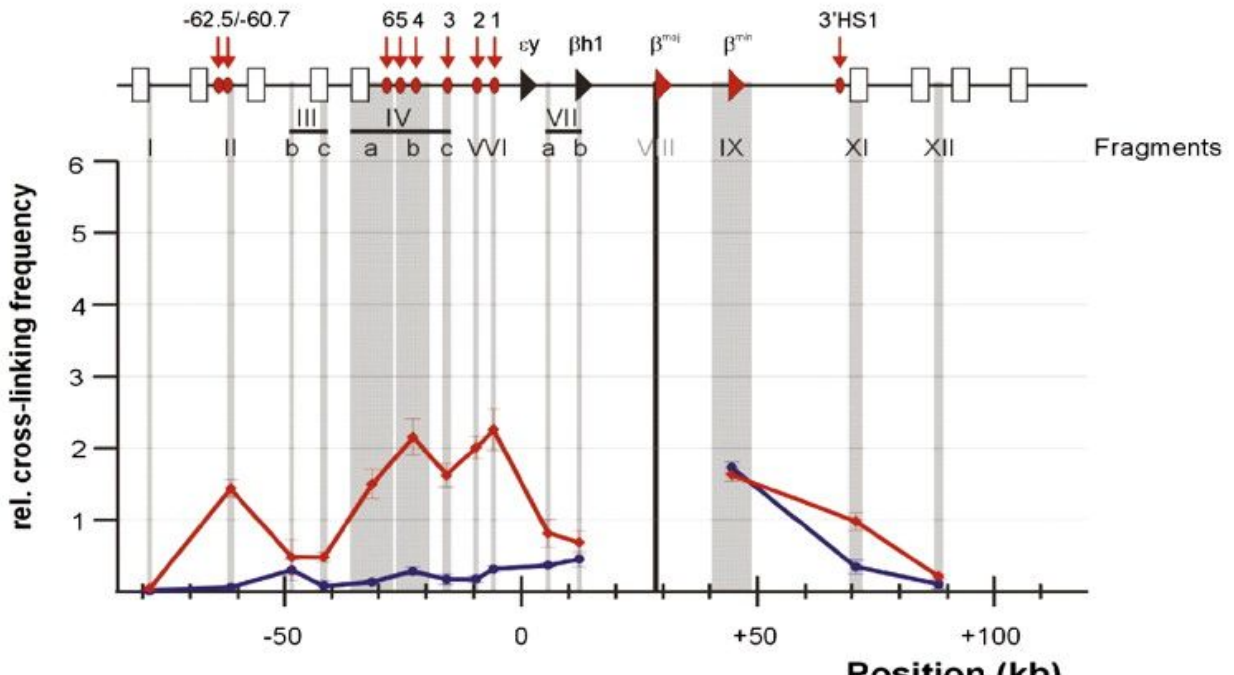


The peak results for the H3K4me3 target compare favorably with the ENCODE annotation data for this well characterized target, with a representative peak for the histone mark target H3K4me3 and a corresponding ENCODE reference peak.

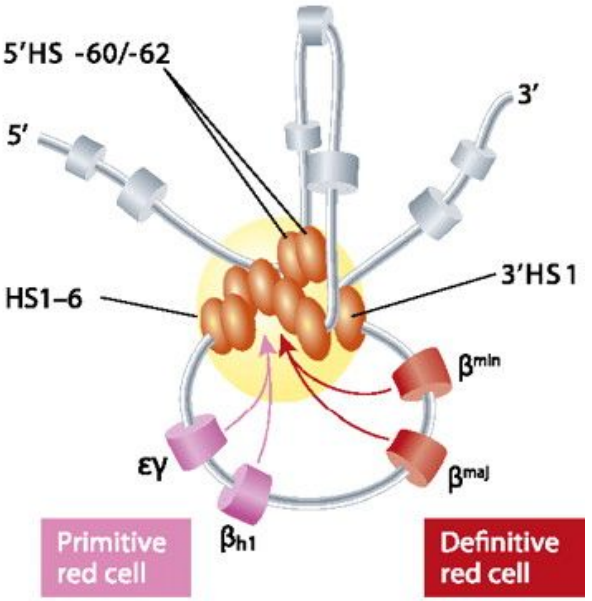
CHROMATIN CONFORMATION CAPTURE (HI-C/3C-SEQ)

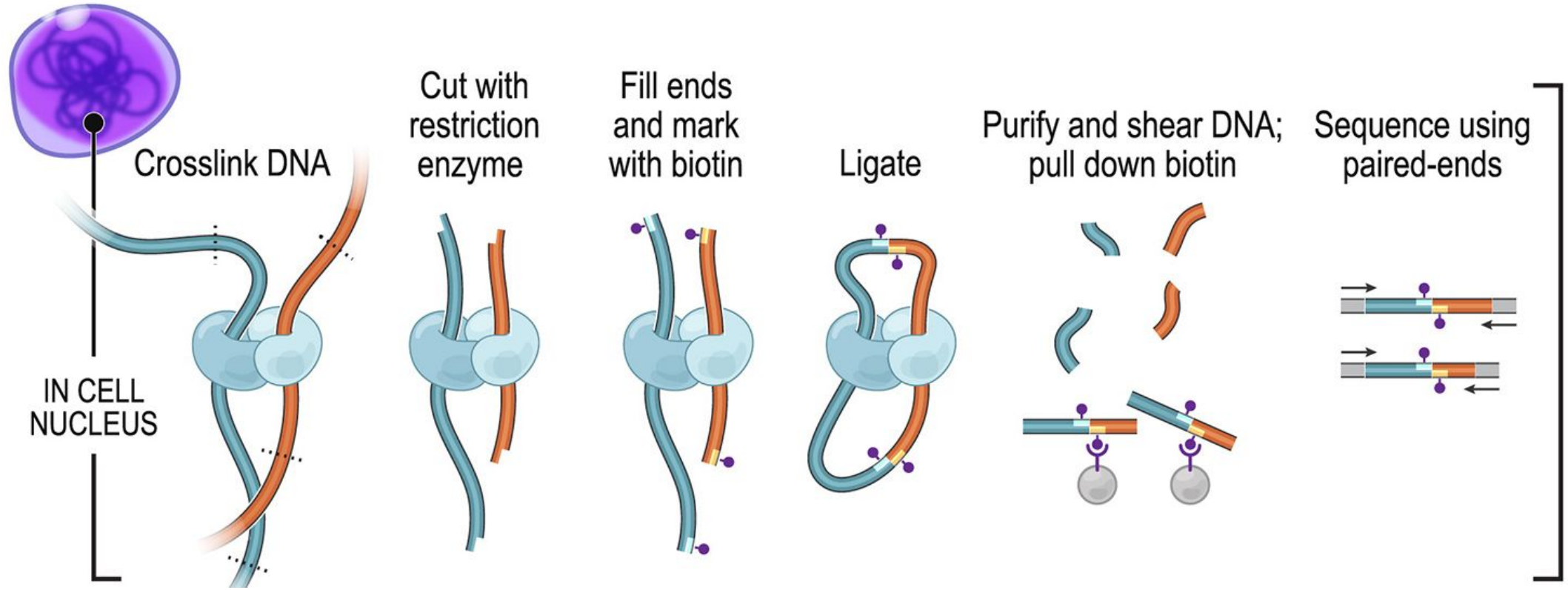


A



B




Crosslink DNA
IN CELL NUCLEUS
Cut with restriction enzyme
Fill ends and mark with biotin
Ligate
Purify and shear DNA; pull down biotin
Sequence using paired-ends

Methods. 2012 Nov;58(3):268-76. doi: 10.1016/j.ymeth.2012.05.001. Epub 2012 May 29.

Hi-C: a comprehensive technique to capture the conformation of genomes.

Belton JM[1], McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J.

Author information

1 Programs in Systems Biology and Gene Function and Expression, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA.

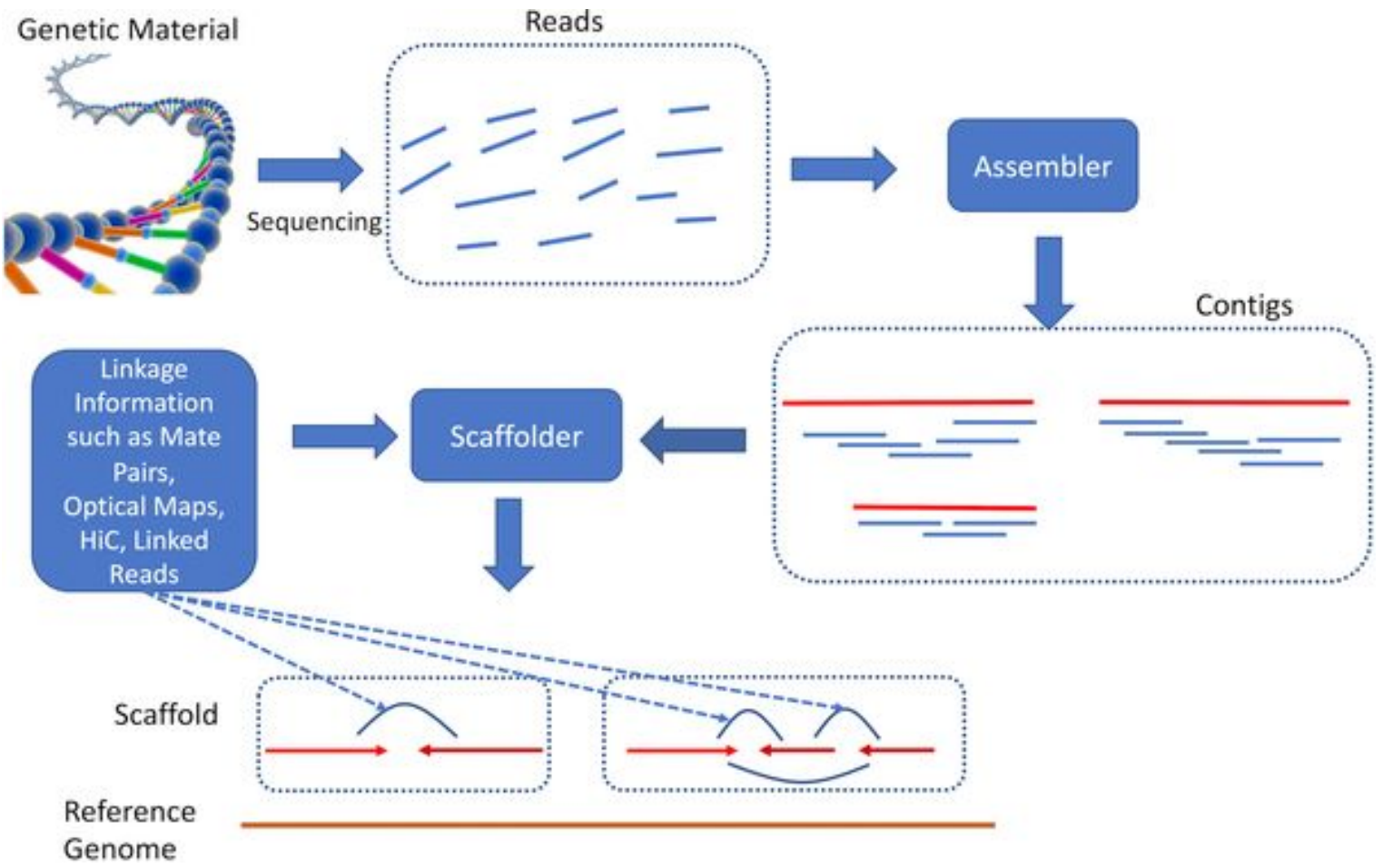
Abstract

We describe a method, Hi-C, to comprehensively detect chromatin interactions in the mammalian nucleus. This method is based on Chromosome Conformation Capture, in which chromatin is crosslinked with formaldehyde, then digested, and re-ligated in such a way that only DNA fragments that are covalently linked together form ligation products. The ligation products contain the information of not only where they originated from in the genomic sequence but also where they reside, physically, in the 3D organization of the genome. In Hi-C, a biotin-labeled nucleotide is incorporated at the ligation junction, enabling selective purification of chimeric DNA ligation junctions followed by deep sequencing. The compatibility of Hi-C with next generation sequencing platforms makes it possible to detect chromatin interactions on an unprecedented scale. This advance gives Hi-C the power to both explore the biophysical properties of chromatin as well as the implications of chromatin structure for the biological functions of the nucleus. A massively parallel survey of chromatin interaction provides the previously missing dimension of spatial context to other genomic studies. This spatial context will provide a new perspective to studies of chromatin and its role in genome regulation in normal conditions and in disease.

Copyright © 2012. Published by Elsevier Inc.

https://www.jove.com/video/1869/hi-c-un-metodo-per-studiare-larchitettura-tridimensionale-dei-genomi?language=Italian

HiC for de novo assembly genome sequences

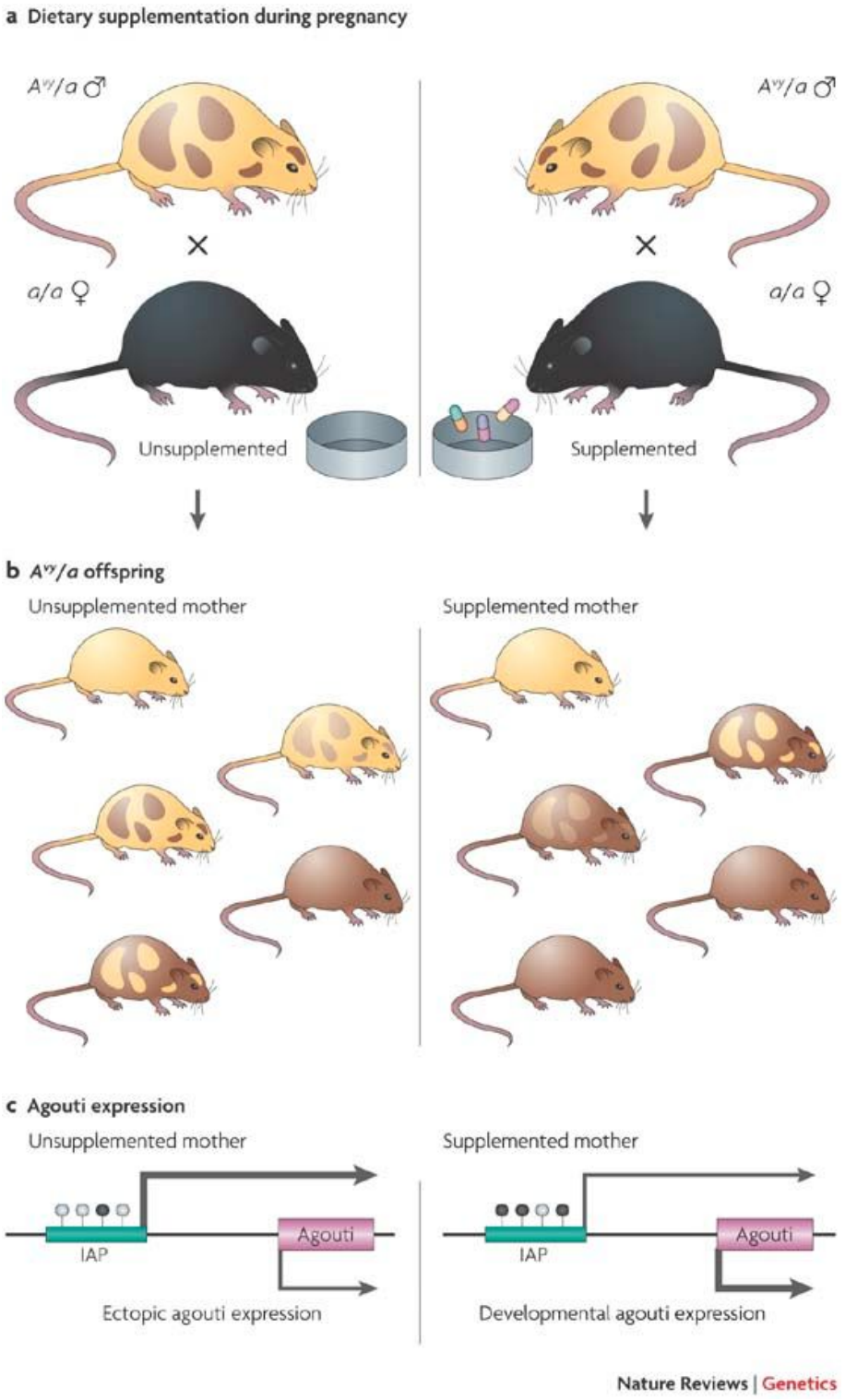


Ghurye J, Pop M (2019) Modern technologies and algorithms for scaffolding assembled genomes. PLOS Computational Biology 15(6): e1006994. https://doi.org/10.1371/journal.pcbi.1006994
https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006994

PLOS | COMPUTATIONAL BIOLOGY

DNA Methylation

- DNA methylation and hydroxymethylation are involved in development, X-chromosome inactivation, cell differentiation, tissue-specific gene expression, plant epigenetic variation, imprinting, cancers, and diseases.
- Methylation usually occurs at the 5’ position of cytosines and plays a crucial role in gene regulation and chromatin remodeling.



- The active agouti gene in mice codes for yellow coat color. When pregnant mice with the active agouti gene are fed a diet rich in methyl donors, the offspring are born with the agouti gene turned off. This effect has been used as an epigenetic biosensor for nutritional and environmental alterations on the fetal epigenome.

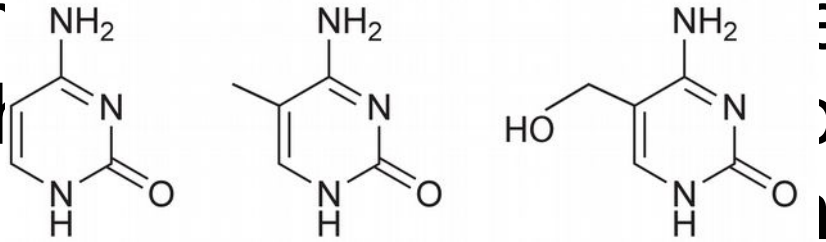
Environmental epigenomics and disease susceptibility

Randy L. Jirtle and Michael K. Skinner

Nature Reviews Genetics 8, 253-262 (April 2007)

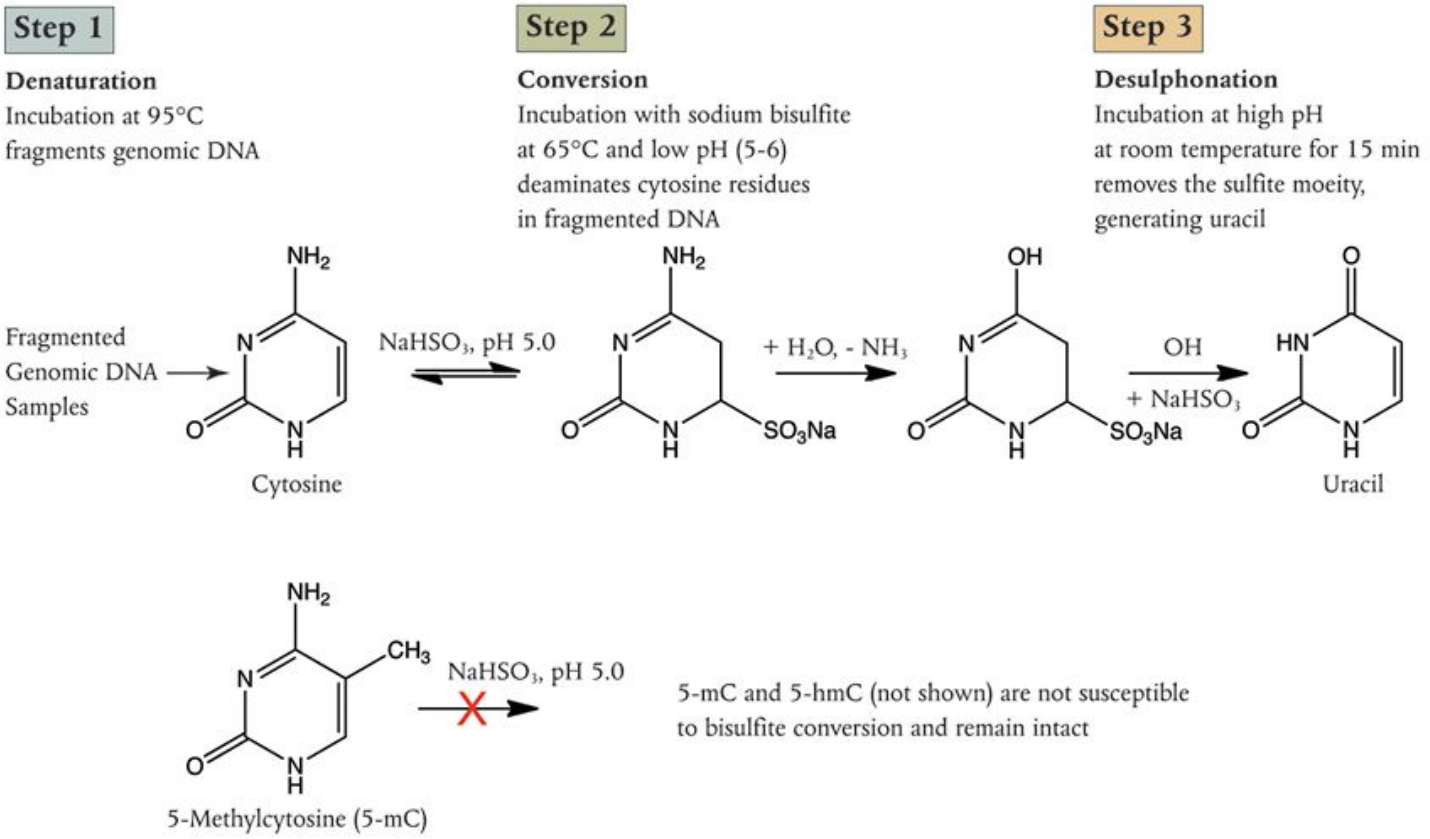
doi:10.1038/nrg2045

- Most cytosine methylation occurs on cytosines located near guanines, called CpG sites. These CpG sites are often located upstream of promoters, or within the gene body. **CpG islands are defined as regions that are greater than 500 bp in length with greater than 55% GC and an expected/observed CpG ratio of > 0.65.**
- While cytosine methylation (5mC) is known as a silencing r enes, cytosine hydroxymeth own to be an activating m e expression and is a pro the DNA



BISULFITE SEQUENCING (BS-SEQ)

- Bisulfite sequencing (BS-Seq) or whole-genome bisulfite sequencing (WGBS) is a well-established protocol to detect methylated cytosines in genomic DNA.
- In this method, genomic DNA is treated with sodium bisulfite and then sequenced, providing single-base resolution of methylated cytosines in the genome. Upon bisulfite treatment, unmethylated cytosines are deaminated to uracils which, upon sequencing, are converted to thymidines.
- Simultaneously, methylated cytosines resist

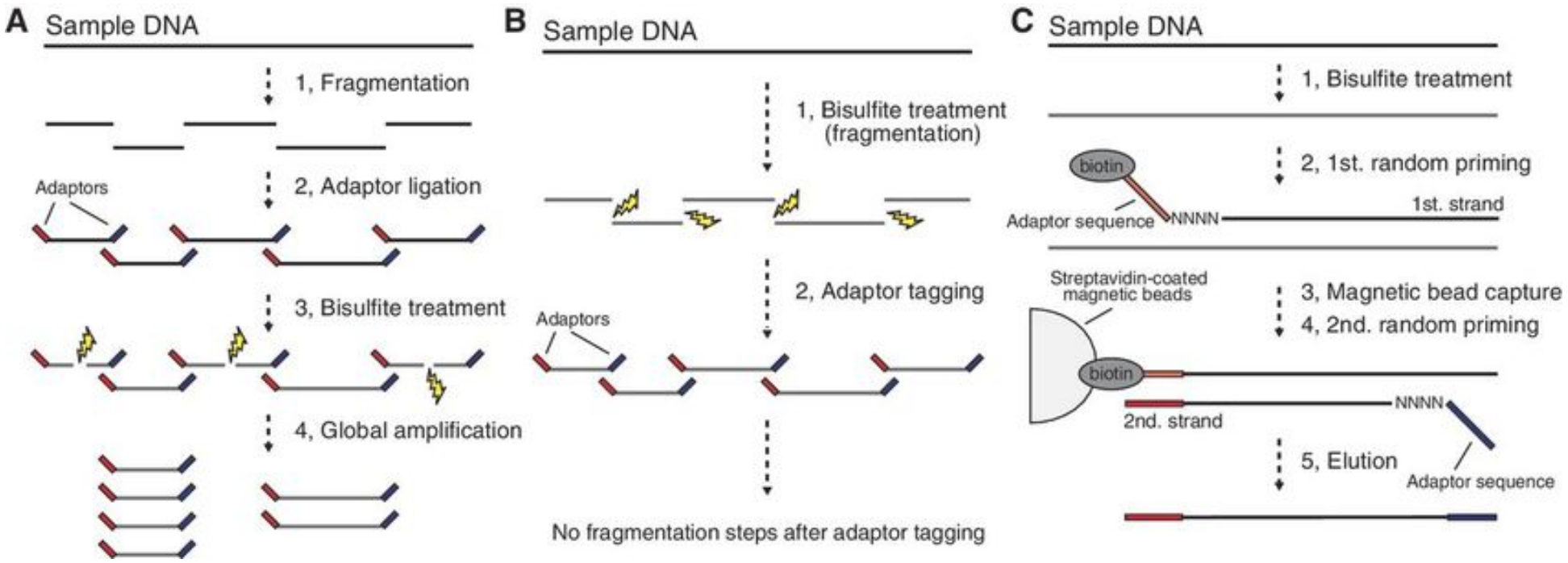


- Bisulfite converts unmethylated cytosines to thymidines, reducing sequence complexity, which can make it difficult to create alignments

- SNPs where a cytosine is converted to thymidine will be missed upon bisulfite conversion

- Bisulfite conversion does not distinguish between 5mC and 5hmC

WGBS PBAT


A
Sample DNA
1, Fragmentation
Adaptors
2, Adaptor ligation
3, Bisulfite treatment
4, Global amplification
B
Sample DNA
1, Bisulfite treatment
(fragmentation)
2, Adaptor tagging
Adaptors
No fragmentation steps after adaptor tagging
C
Sample DNA
1, Bisulfite treatment
biotin
2, 1st. random priming
Adaptor sequence
NNNN
1st. strand
Streptavidin-coated
magnetic beads
3, Magnetic bead capture
4, 2nd. random priming
biotin
NNNN
2nd. strand
5, Elution
Adaptor sequence

Sequence Rearrangements

A growing body of evidence suggests that somatic genomic rearrangements, such as retrotransposition and copy number variants (CNVs), are relatively common in healthy individuals.

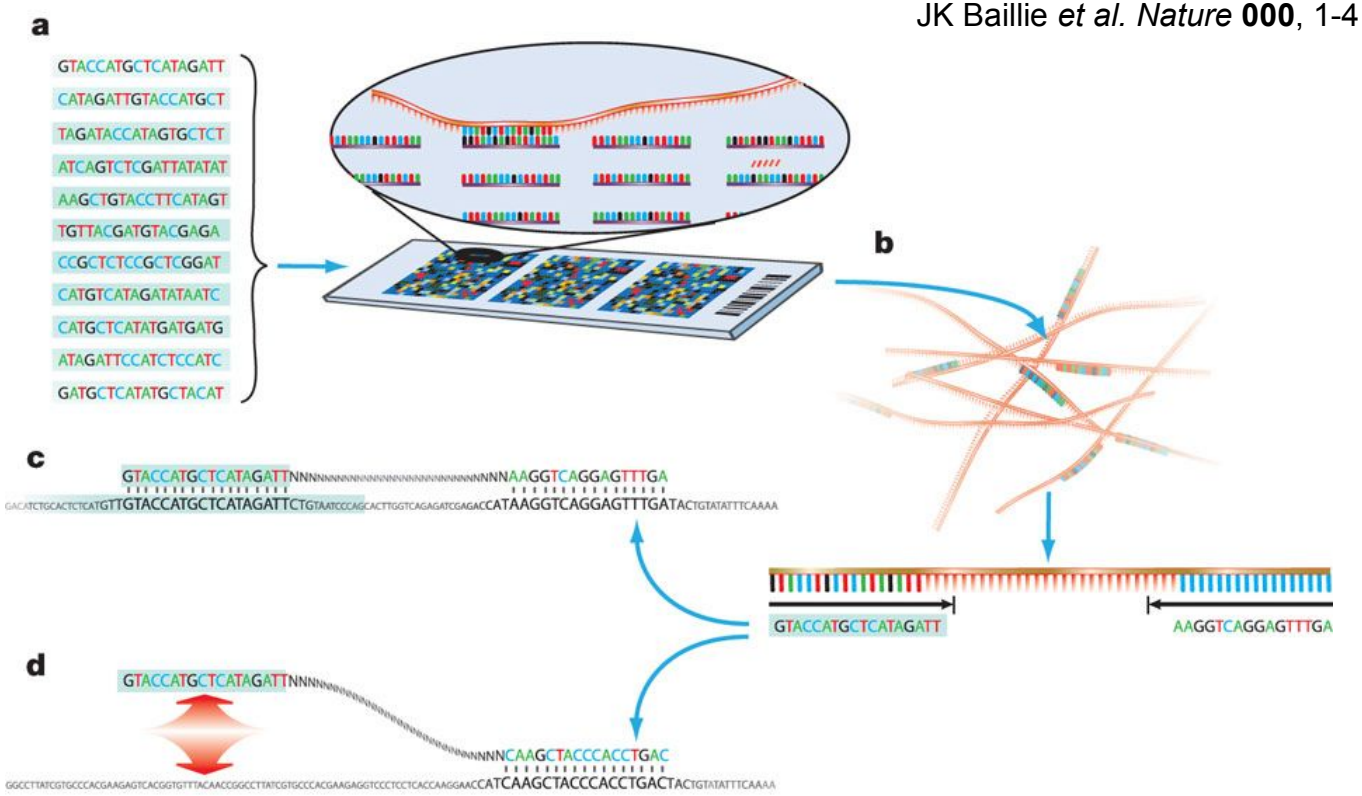
Cancer genomes are also known to contain numerous complex rearrangements. While many of these rearrangements can be detected during routine next-generation sequencing, specific techniques are available to study rearrangements such as transposable elements.

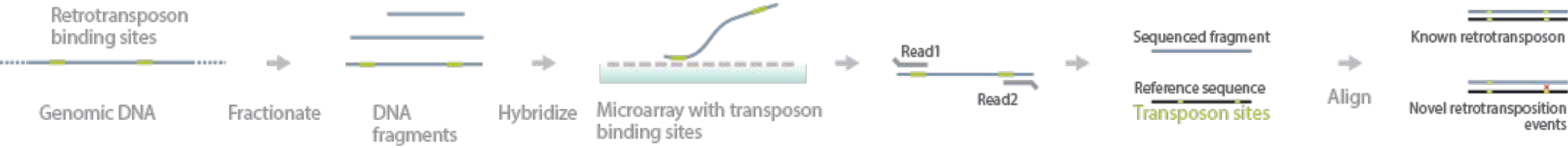
Transposable genetic elements (TEs) comprise a vast array of DNA sequences with the ability to move to new sites in genomes either directly by a cut-and-paste mechanism (transposons) or indirectly through an RNA intermediate (retrotransposons).

TEs make up about 66-69% of the human genome and play roles in ageing, cancers, brain, development, embryogenesis, and phenotypic variation in populations and evolution. TEs played a major role in dynamic

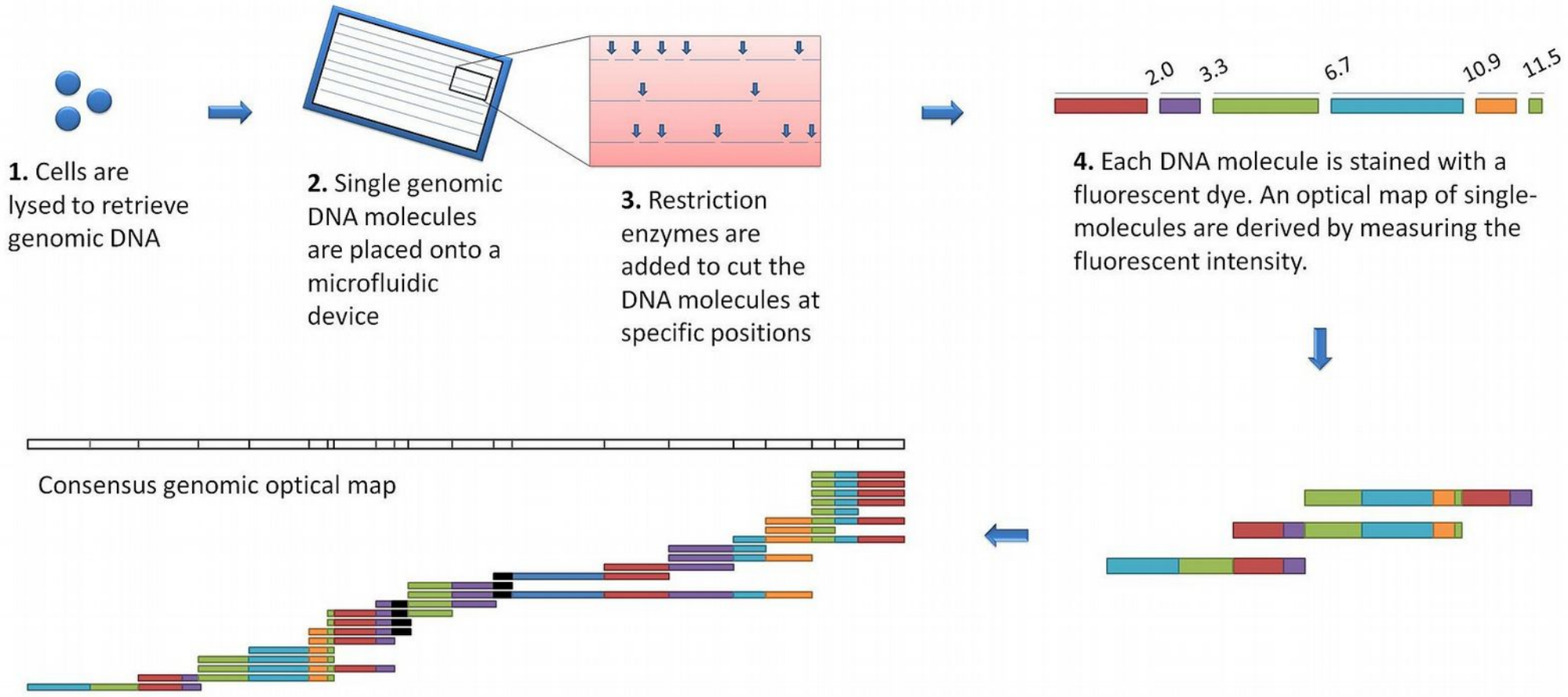
Overall RC-seq methodology.

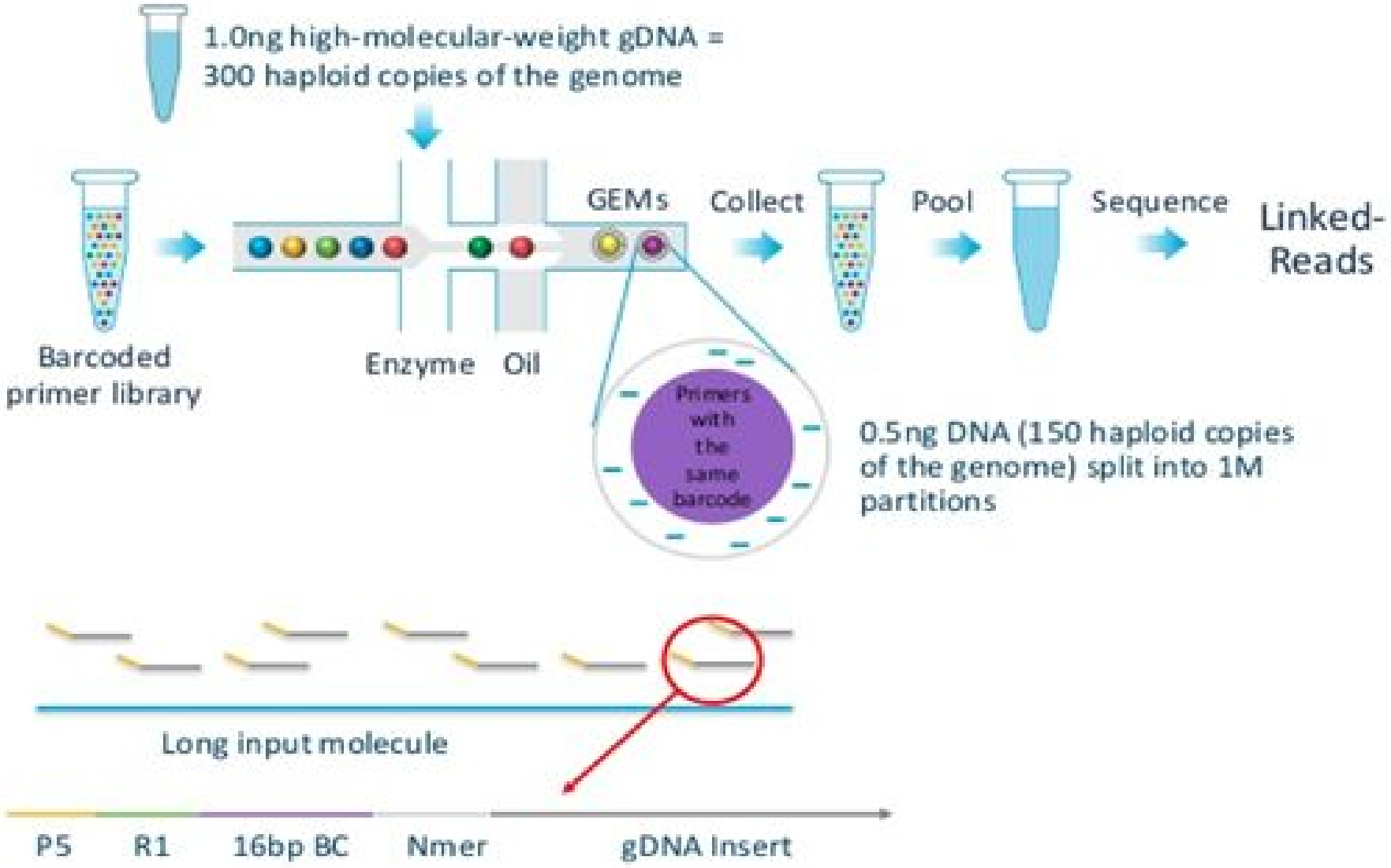
JK Baillie *et al. Nature* **000**, 1-4 (2011) doi:10.1038/nature10531





Human genome resequencing. Optical mapping




1.0ng high-molecular-weight gDNA =
300 haploid copies of the genome
Barcoded
primer library
Enzyme
Oil
GEMs
Collect
Pool
Sequence
Linked-
Reads
Primers
with
the
same
barcode
0.5ng DNA (150 haploid copies
of the genome) split into 1M
partitions
Long input molecule
P5
R1
16bp BC
Nmer
gDNA Insert

10X genomics

https://www.youtube.com/watch?v=XwBI13Q4ilo

10X genomics

SINGLE CELL TRANSCRIPTOMICS

Single Cell Gene Expression

Go beyond traditional gene expression analysis to characterize cell populations, cell types, cell states, and more

Single Cell Immune Profiling

Simultaneously examine the cellular context of the adaptive immune response in hundreds to tens of thousands of lymphocytes

SINGLE CELL GENOMICS

Single Cell CNV

Reveal genome heterogeneity and understand clonal evolution

SINGLE CELL EPIGENOMICS

Single Cell ATAC

Reveal chromatin accessibility cell by cell to decipher gene regulatory mechanisms

LINKED-READS GENOMICS

Genome Sequencing

Resolve phasing, structural variants, and variants in previously inaccessible parts of the genome

Exome Sequencing

Resolve genic phasing, structural variation, and variants in previously inaccessible and complex regions

de novo Assembly

Enable true diploid genome assembly like never before

SPATIAL GENE EXPRESSION

Spatial Gene Expression

Understand gene expression levels with morphological context by performing mRNA analysis on intact tissue