

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

CLASH-SEQ

TIF-SEQ/PEAT

IN-SEQ

TAB-SEQ

ICE
OS-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

MBDCAP-SEQ

HITS-CLIP

PARE-SEQ/GMUT

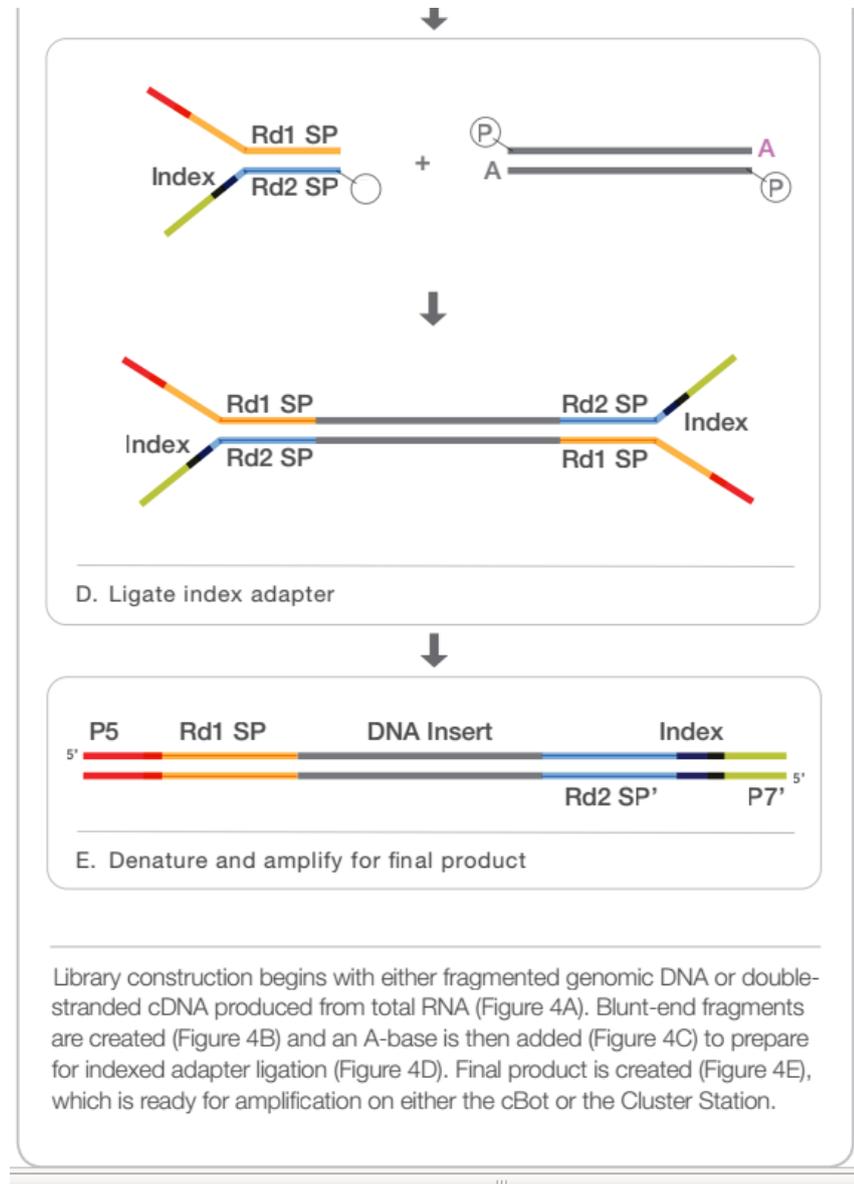
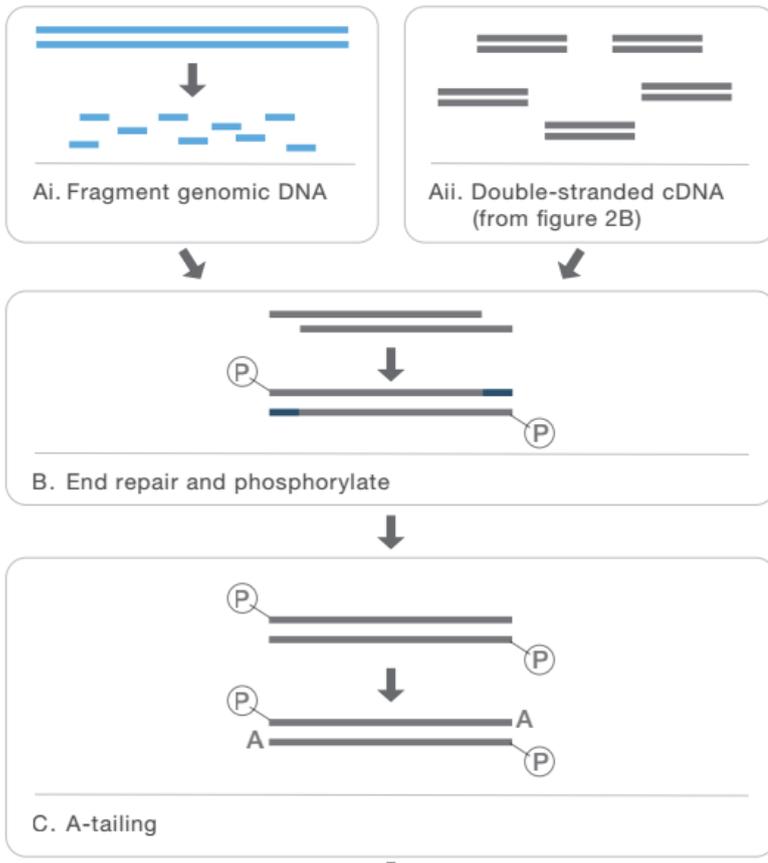
HI-C/3-C

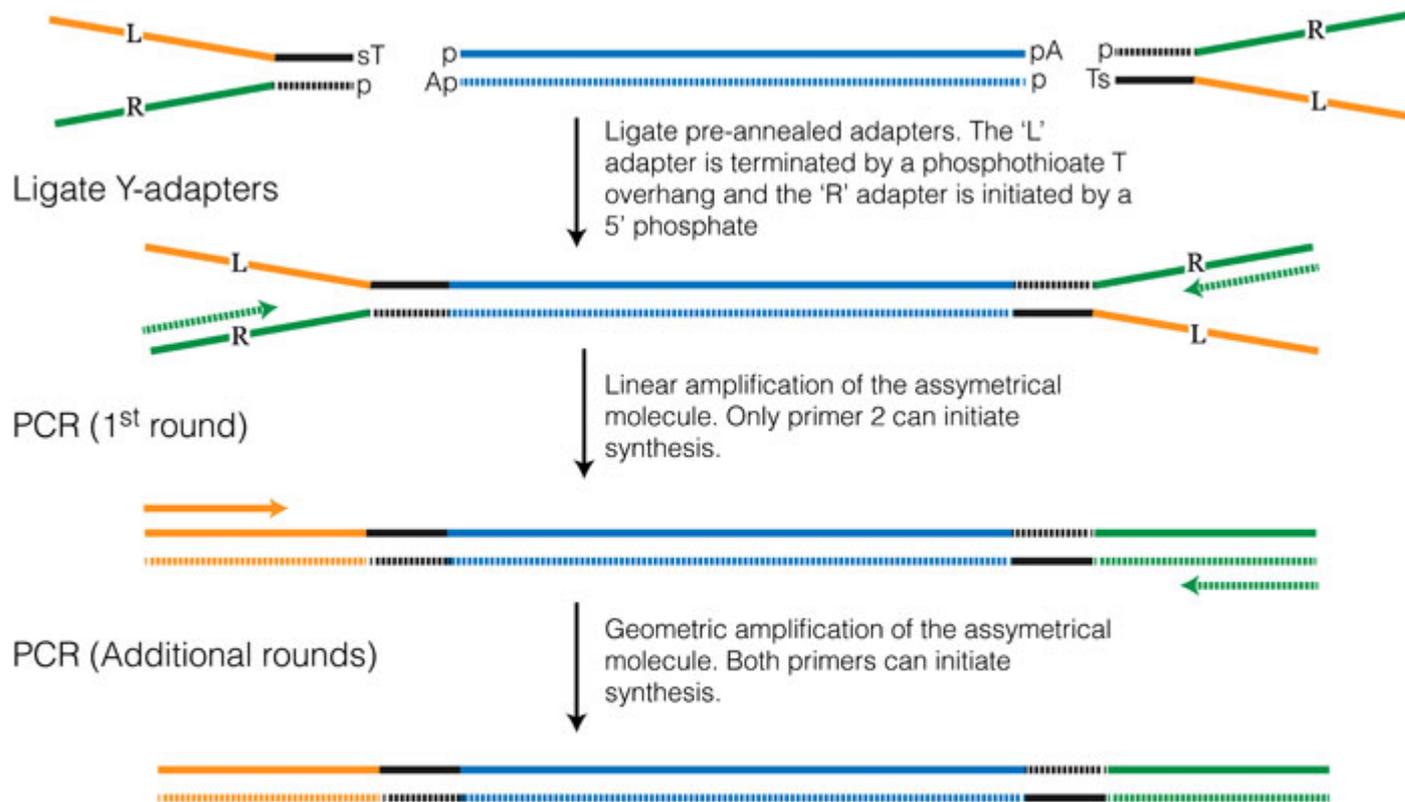
TRAP-SEQ
RC-SEQ
FRAG-SEQ

- 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

Figure 4: Adapter Ligation Results in Sequence-Ready Constructs without PCR





Ligate pre-annealed adapters. The 'L' adapter is terminated by a phosphothioate T overhang and the 'R' adapter is initiated by a 5' phosphate

Linear amplification of the assymetrical molecule. Only primer 2 can initiate synthesis.

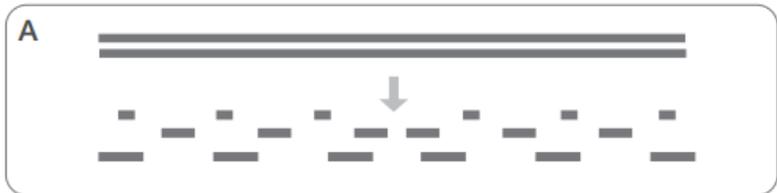
Geometric amplification of the assymetrical molecule. Both primers can initiate synthesis.

Primers

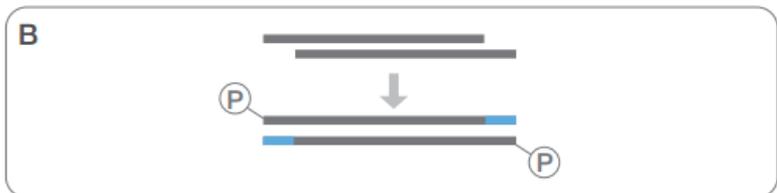
-  Primer 1: Equivalent to the forward strand of sequence L
-  Primer 2: Equivalent to the complementary strand of sequence R

TruSeq® DNA PCR-Free Sample Preparation Kit

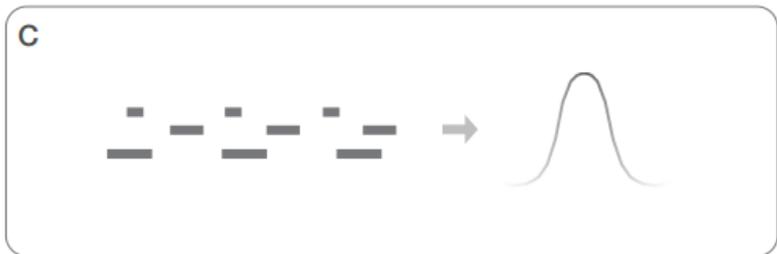
Figure 2: Adapter Ligation Results in Sequence-Ready Constructs without PCR



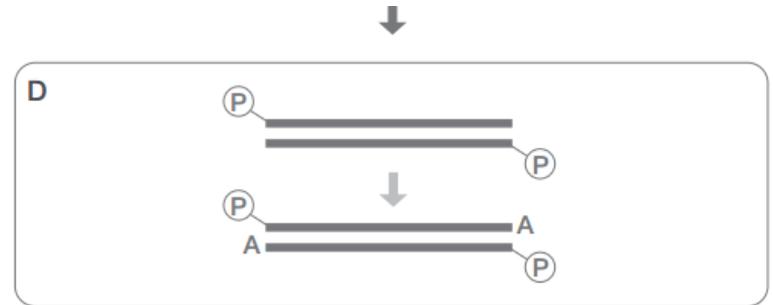
Library construction begins with genomic DNA that is subsequently fragmented



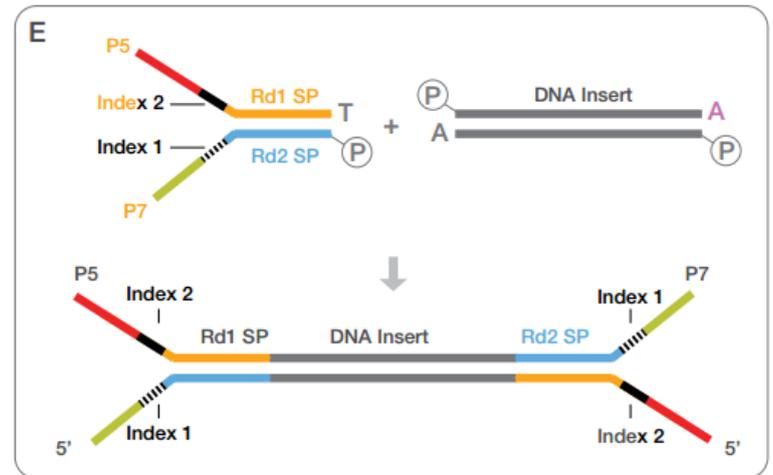
Blunt-end fragments are created.



Fragments are narrowly size selected with sample purification beads.



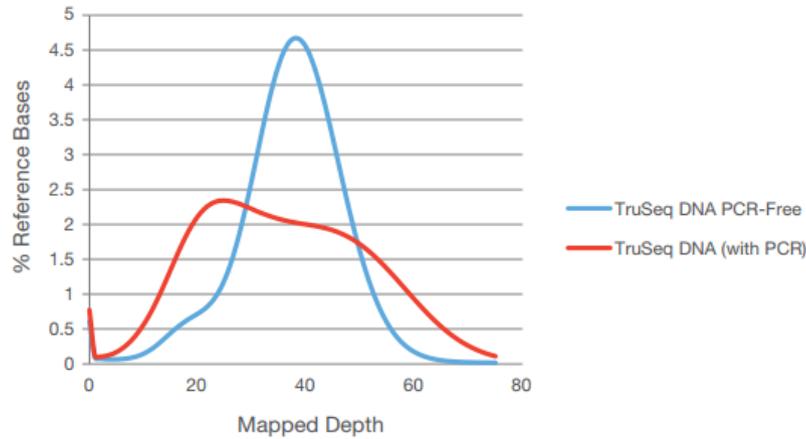
A-base is added.



Dual-index adapters are ligated to the fragments* and final product is ready for cluster generation.

*The TruSeq DNA PCR-Free LT indexing solution features a single-index adapter at this step.

Figure 4: Greater Coverage Uniformity

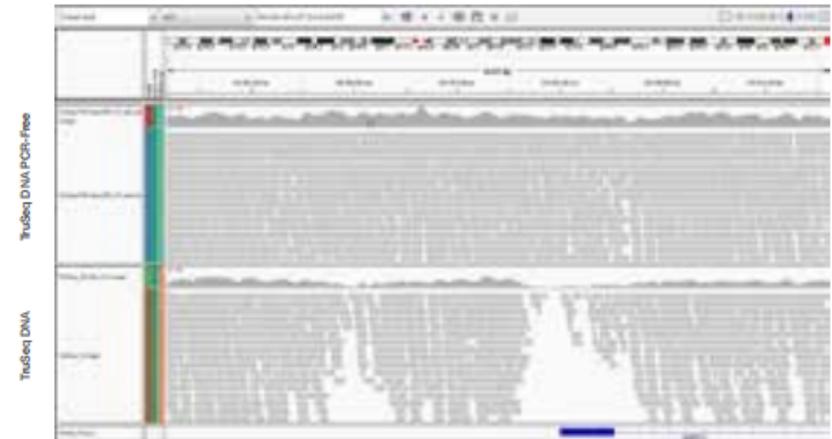


TruSeq DNA PCR-Free libraries provide greater coverage uniformity across the genome when compared to those generated using the TruSeq DNA protocol.

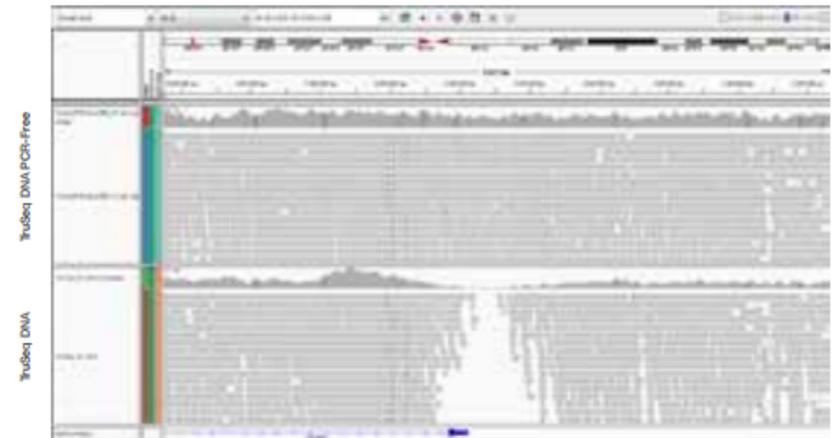
The PCR-Free kit also provides superior coverage of traditionally challenging genomic content, including GC-rich regions, promoters, and repetitive regions (Figure 5), allowing researchers to access more genomic information from each sequencing run (Figure 6).

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

A



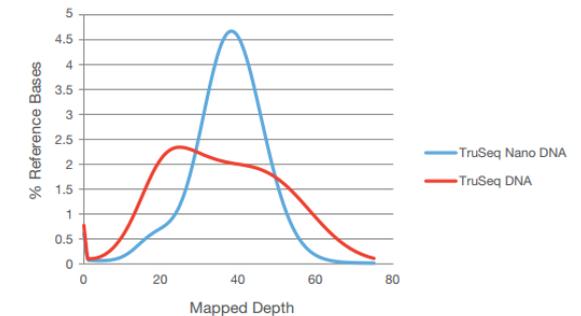
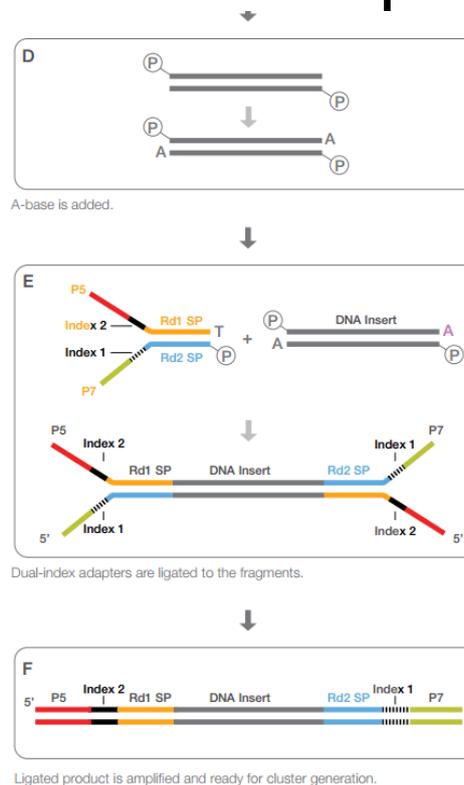
B



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.



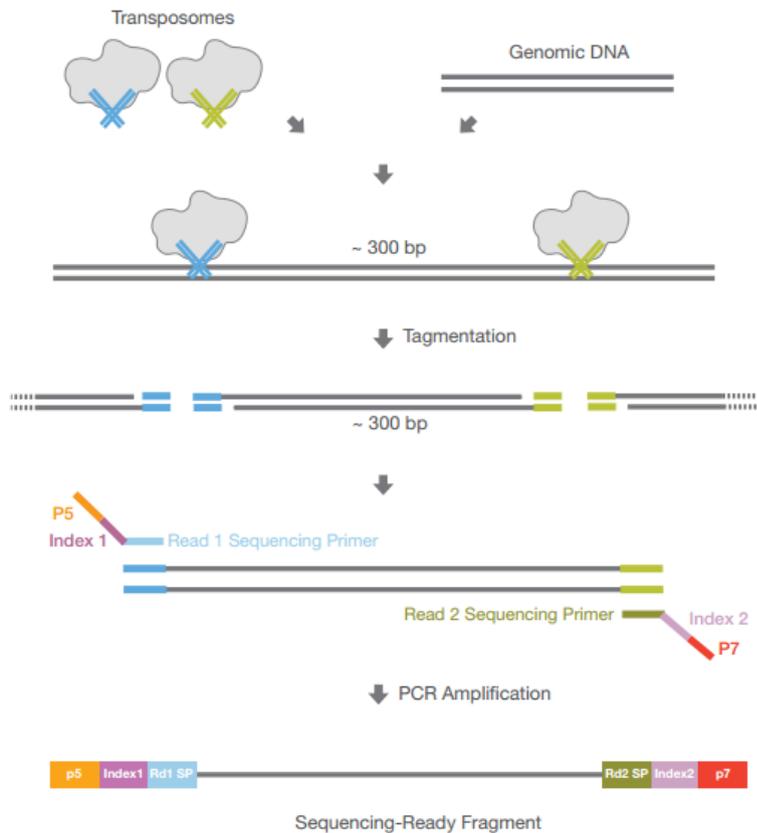
TruSeq Nano DNA libraries provide greater coverage uniformity across the genome when compared to those generated using the TruSeq DNA protocol.

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

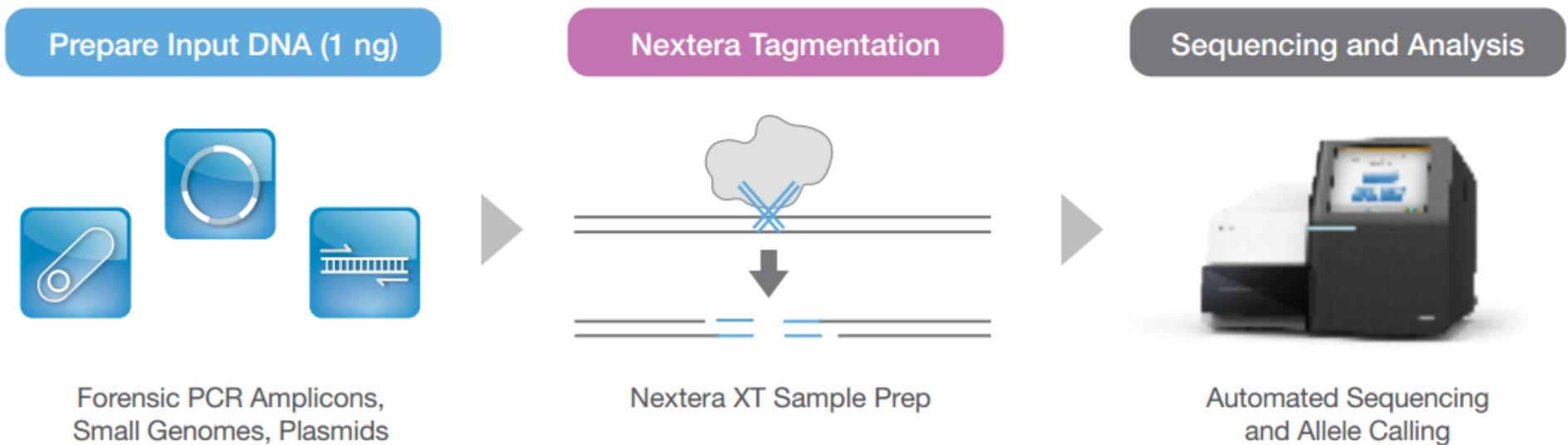
- Large-genome resequencing
- Small-genome resequencing
- Amplicon resequencing
- Clone or plasmid sequencing

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.

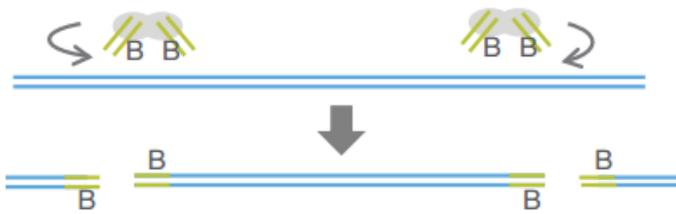
Nextera XT

Figure 1: Nextera XT Sample Preparation Workflow

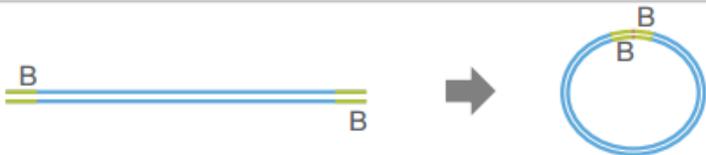


The combination of Nextera XT and rapid sequencing with the MiSeq System provides a complete DNA to data workflow in only 8 hours.

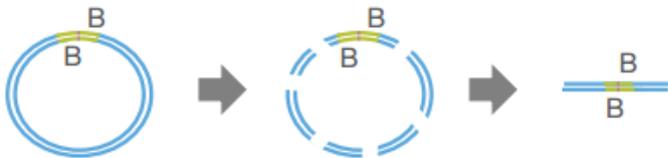
Figure 1: Nextera Mate Pair Workflow



Genomic DNA (blue) is tagmented with a Mate Pair Tagment Enzyme, which attaches a biotinylated junction adapter (green) to both ends of the tagmented molecule.



The tagmented DNA molecules are then circularized and the ends of the genomic fragment are linked by two copies of the biotin junction adapter.

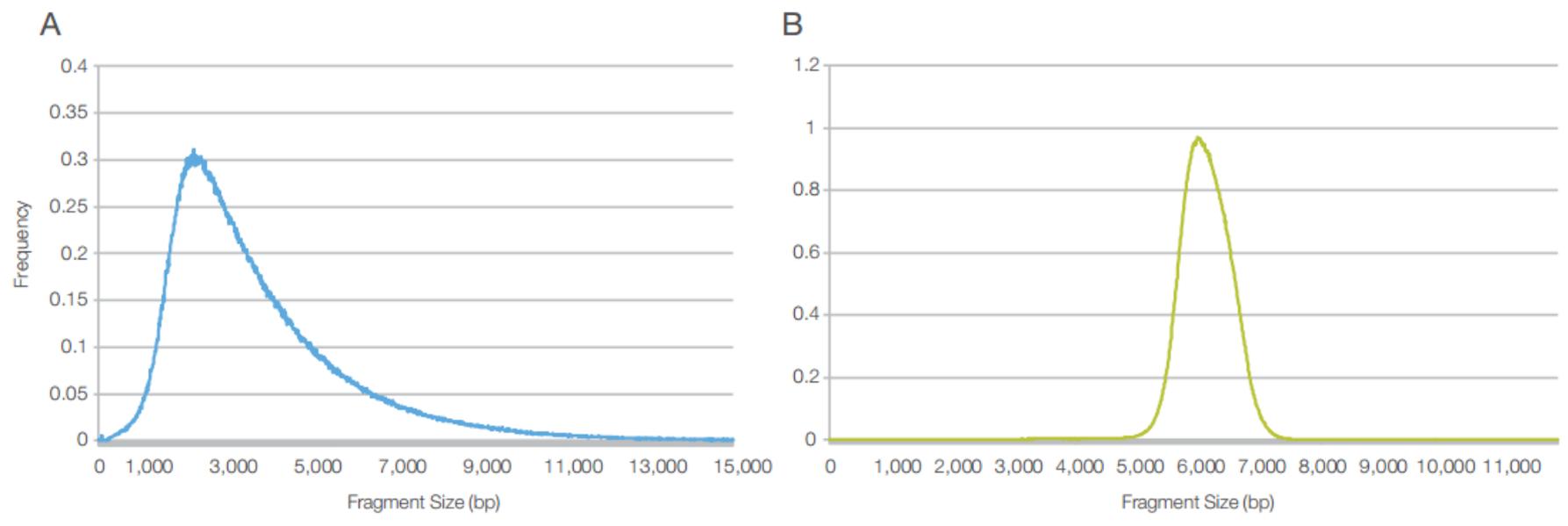


Circularized molecules are then fragmented again, yielding smaller fragments. Sub-fragments containing the original junction are enriched via the biotin tag (B) in the junction adapter.



After End Repair and A-Tailing, TruSeq DNA adapters (gray and purple) are then added, enabling amplification and sequencing.

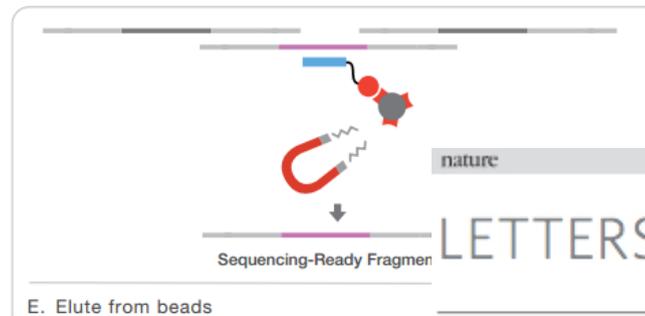
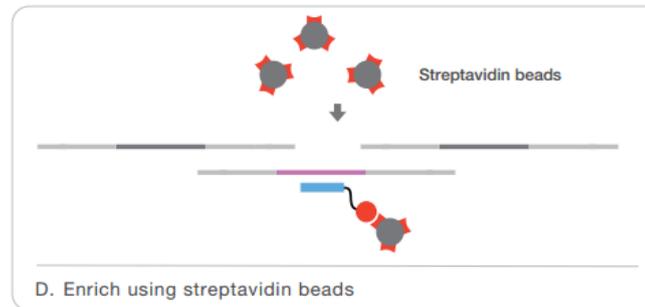
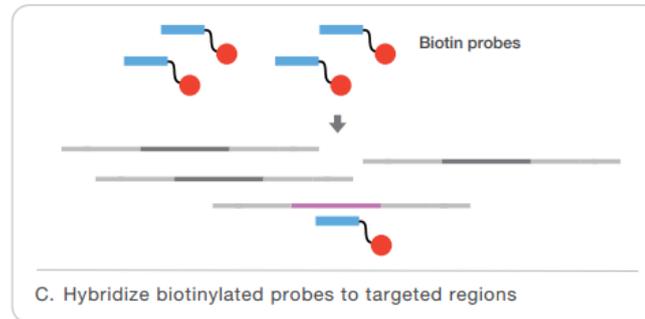
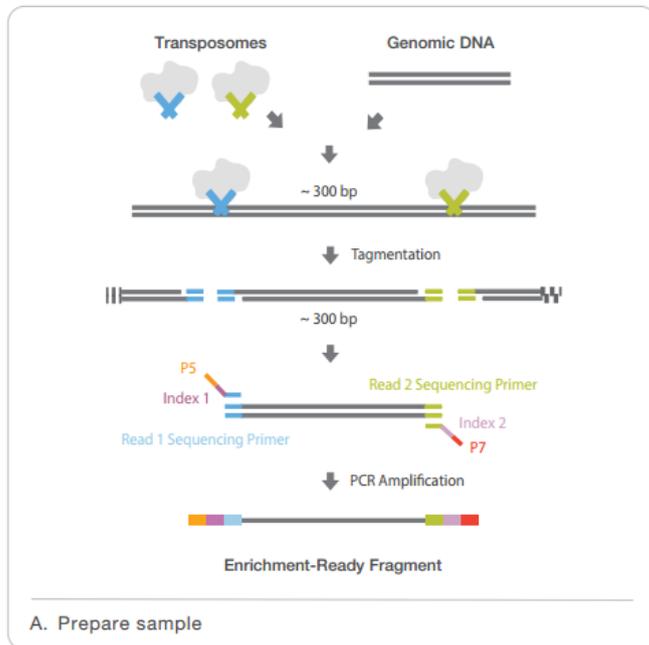
Figure 2: Fragment Size Distribution with Dual Protocols



Panel A shows the fragment size distribution of an *E. coli* mate pair library prepared using the Nextera Mate Pair gel-free protocol, resulting in a broad fragment size distribution. Panel B shows the narrow fragment size distribution of an *E. coli* mate pair library generated with the Nextera Mate Pair gel-plus protocol with automated size selection using the Pippin Prep platform.

Rapid Capture Exomes

Figure 1: Nextera Rapid Capture Workflow



nature

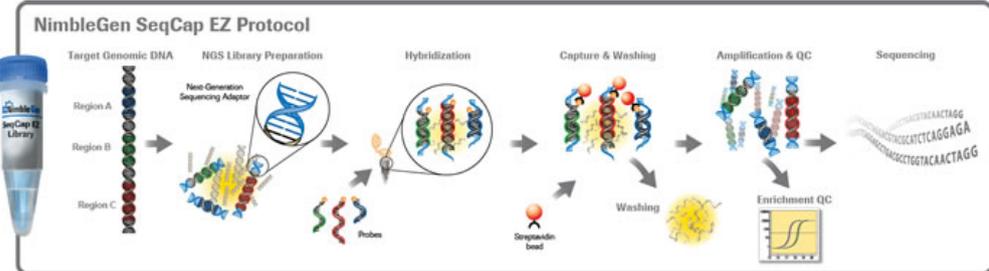
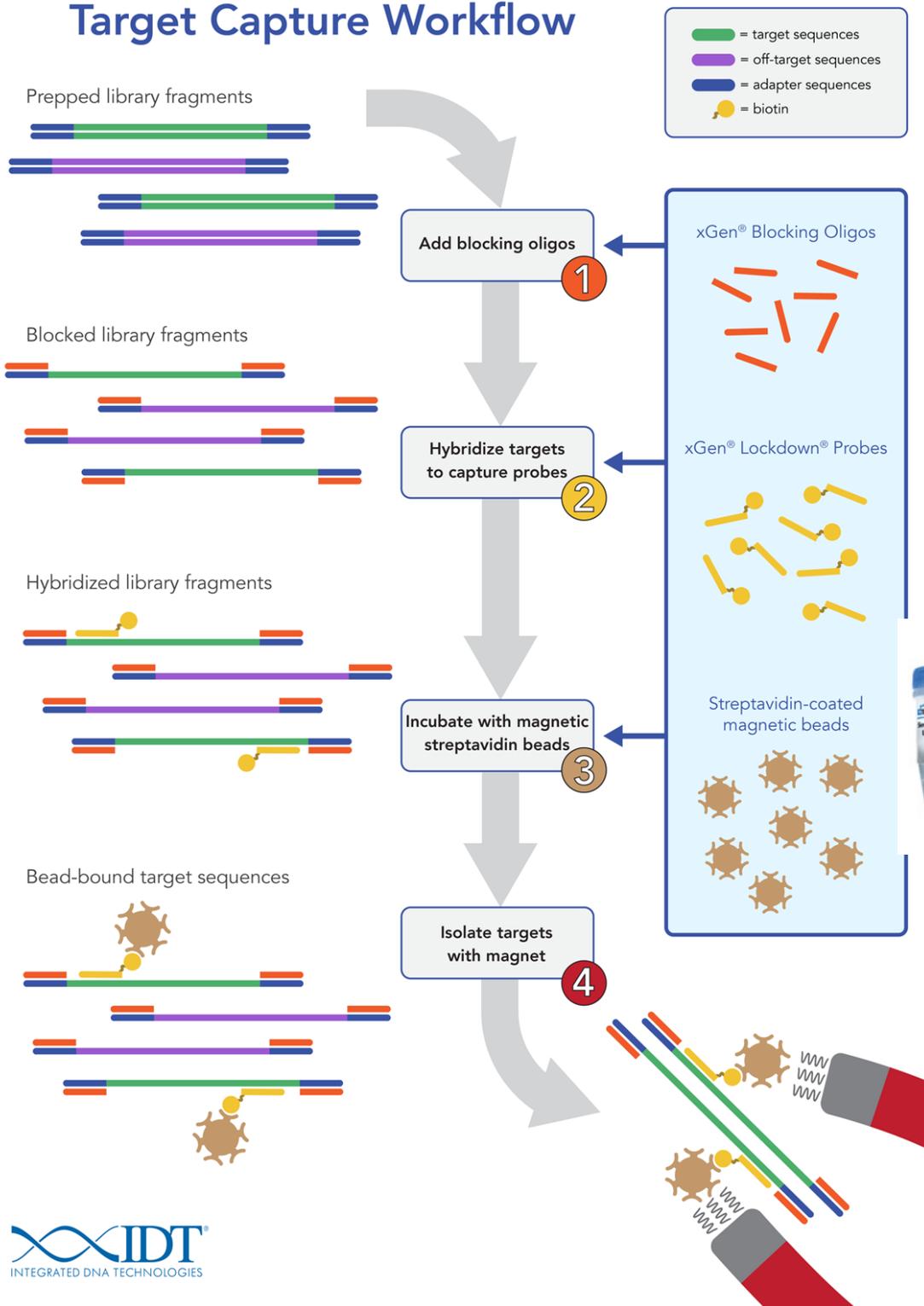
Vol 461 | 10 September 2009 | doi:10.1038/nature08250

LETTERS

Targeted capture and massively parallel sequencing of 12 human exomes

Sarah B. Ng¹, Emily H. Turner¹, Peggy D. Robertson¹, Steven D. Flygare¹, Abigail W. Bigham², Choli Lee¹, Tristan Shaffer¹, Michelle Wong¹, Arindam Bhattacharjee⁴, Evan E. Eichler^{1,3}, Michael Bamshad², Deborah A. Nickerson¹ & Jay Shendure¹

Target Capture Workflow



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.

Ampliseq

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

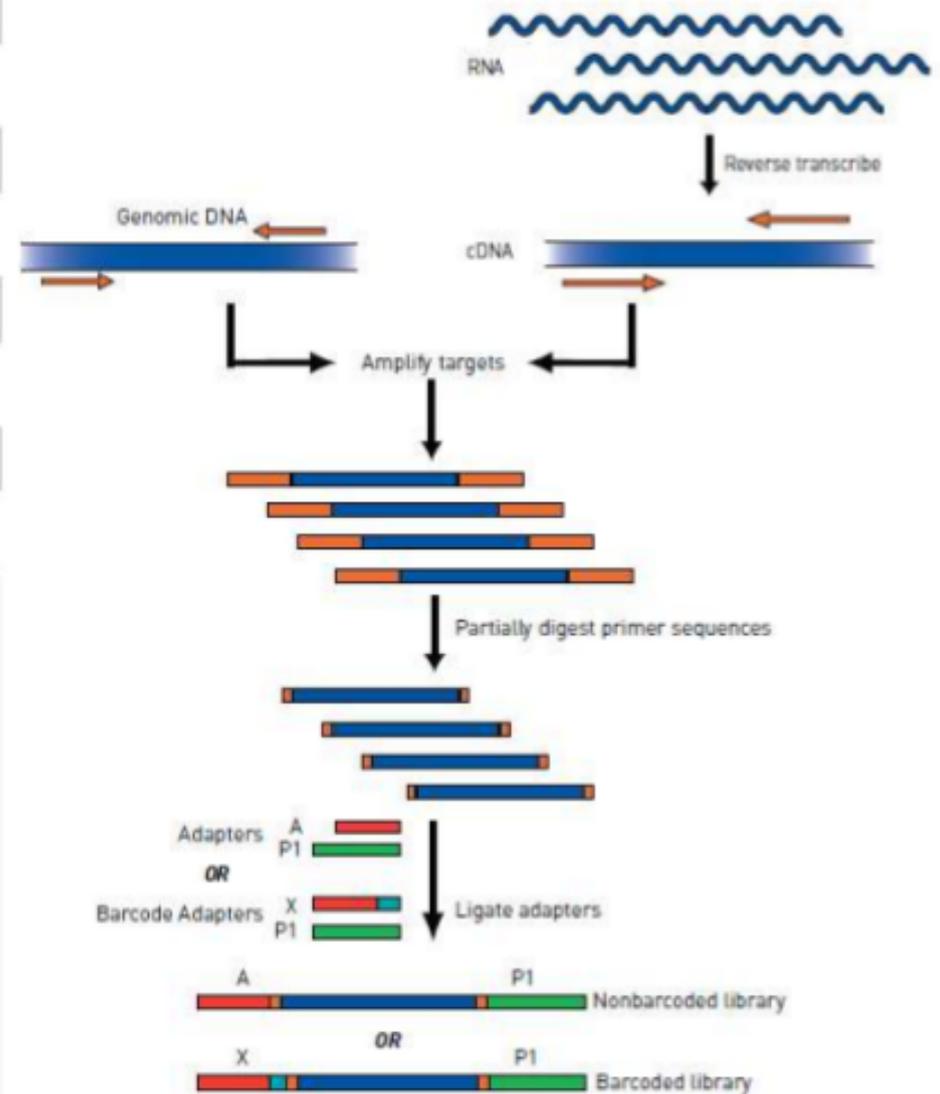
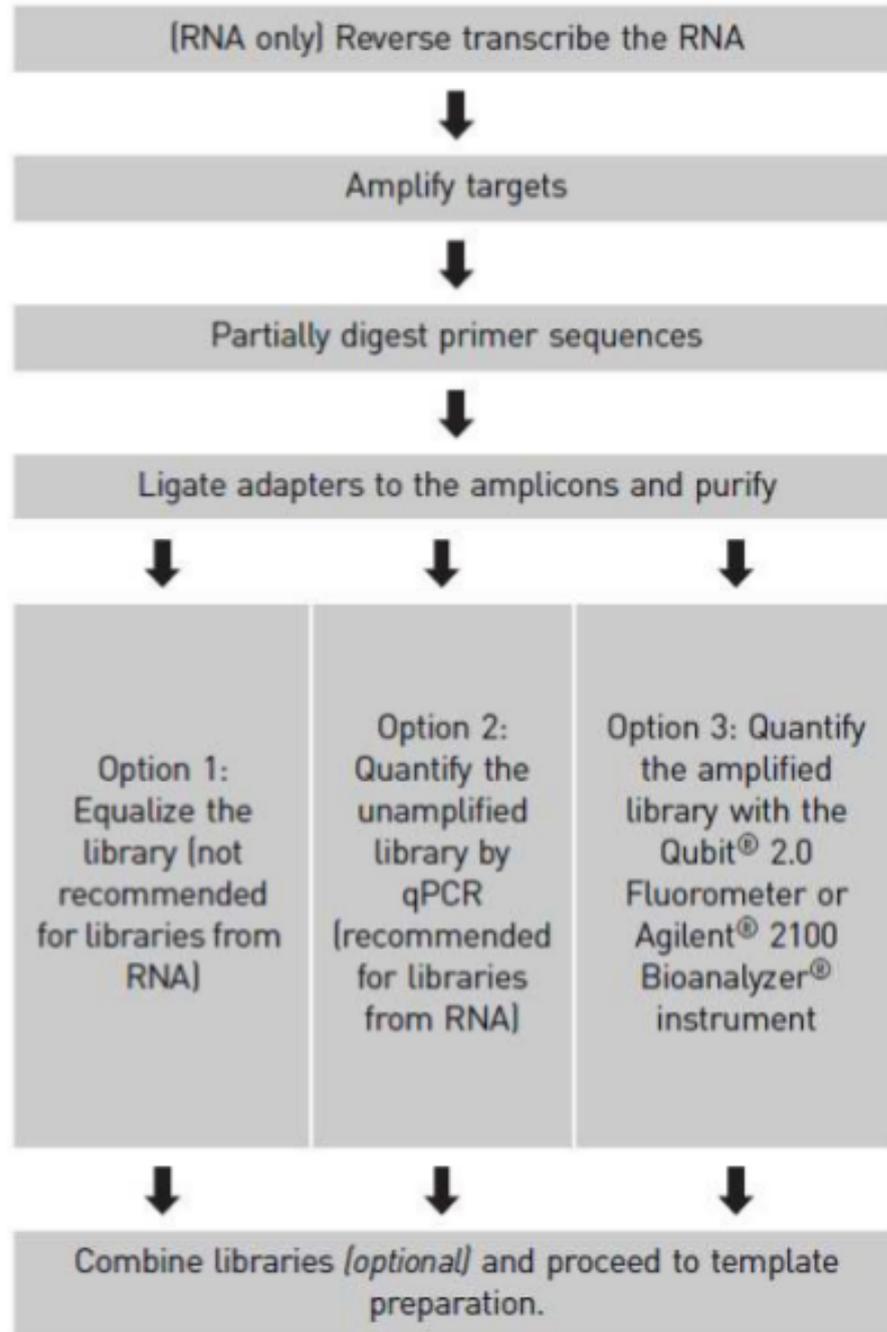
Now, with AmpliSeq™, it is possible to simultaneously amplify, sequence and genotype hundreds of genomic regions in a single project.

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.

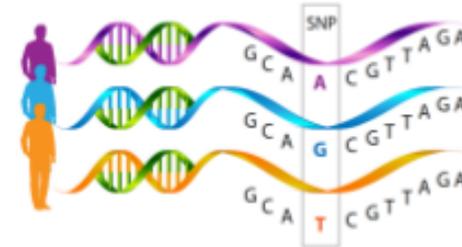
video

Ion AmpliSeq™ Target Selection Overview



Key technology features:

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



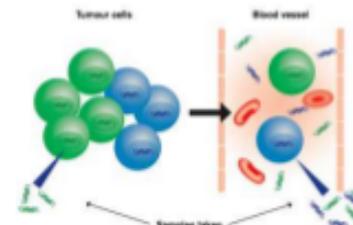
- Designs for “Any Genome”
 - AgBio applications and more



- 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

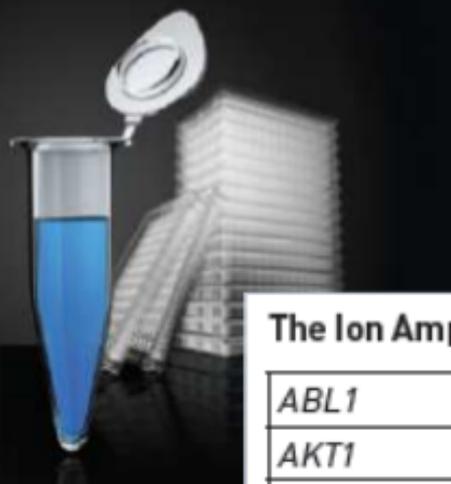
Ready-to-use Human Identity Panel	Ready-to-use Human Ancestry Panel	Ready-to-use RNA Apoptosis Panel	Ready-to-use RNA Cancer Panel
Ready-to-use Cancer Hotspot Panel	Ready-to-use Comprehensive Cancer Panel	Ready-to-use Inherited Disease Panel	Community RNA Fusion Lung Cancer Research Panel
Community BRCA 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

ion torrent
by life technologies

Ion AmpliSeq™ Cancer Hotspot Panel v2

Just one tube. Just 10 ng of DNA. Just one day.



As little as 10ng input
50 genes
207 amplicons

The Ion AmpliSeq™ Cancer Panel targets 50 genes

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

Ion AmpliSeq™ Comprehensive Cancer Panel (CCP)

ion torrent
 by life technologies™

Ion AmpliSeq™ Comprehensive Cancer Panel

Extensive survey of over 400 genes with only 40 ng of DNA

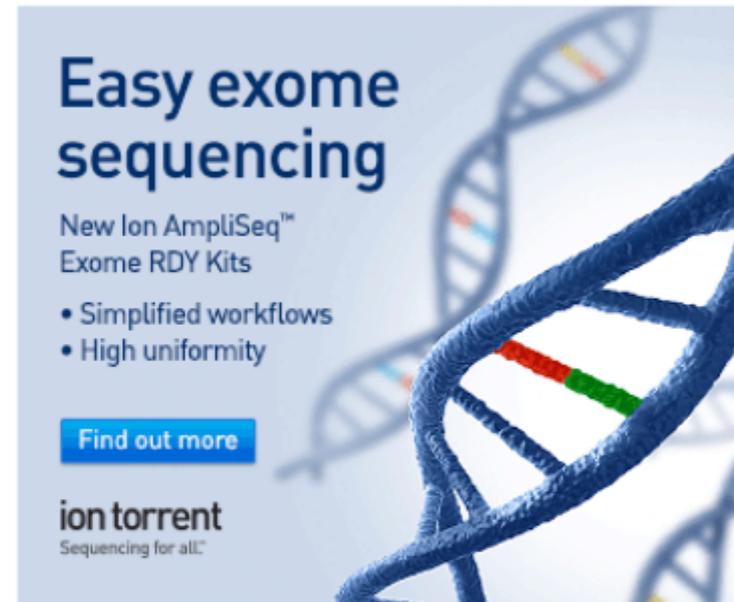


- Targets coding exons in 409 human oncogenes and tumor suppressor genes
- ~16,000 amplicons
- Detection of known COSMIC somatic mutations

Gene	Gene	Gene	Gene	Gene	Gene	Gene	Gene
ABL1	AURKA	BRN3A	CDK2	CTNNA1	EPAS1	FANCD2	FZR1
ABL2	AURKB	BRN3B	CDK4	CYLD	EPAS2	FANCD3	GAPD
ACR2A	AURKC	BRN3C	CDK9	CYP21A	EPAS3	FANCF	GATA1
ADAMT2B	BRN3D	BRN3E	CDKN2A	CYP24A	EPAS4	FANCG	GATA2
AF11	BRN3F	BRN3G	CDKN2B	DAPK	EPAS5	FANCI	GATA3
AF12	BRN3H	BRN3I	CDKN2C	DDIT	EPAS6	FANCD1	BRN4
AF13	BRN3J	BRN3K	CDKN2D	DDIT2	EPAS7	FANCD2	BRN5
AF14	BRN3L	BRN3M	CDKN2E	DDIT3	EPAS8	FANCD3	BRN6
AF15	BRN3N	BRN3O	CDKN2F	DDIT4	EPAS9	FANCD4	BRN7
AF16	BRN3P	BRN3Q	CDKN2G	DDIT5	EPAS10	FANCD5	BRN8
AF17	BRN3R	BRN3S	CDKN2H	DDIT6	EPAS11	FANCD6	BRN9
AF18	BRN3T	BRN3U	CDKN2I	DDIT7	EPAS12	FANCD7	BRN10
AF19	BRN3V	BRN3W	CDKN2J	DDIT8	EPAS13	FANCD8	BRN11
AF20	BRN3X	BRN3Y	CDKN2K	DDIT9	EPAS14	FANCD9	BRN12
AF21	BRN3Z	BRN3AA	CDKN2L	DDIT10	EPAS15	FANCD10	BRN13
AF22	BRN3AB	BRN3AC	CDKN2M	DDIT11	EPAS16	FANCD11	BRN14
AF23	BRN3AD	BRN3AE	CDKN2N	DDIT12	EPAS17	FANCD12	BRN15
AF24	BRN3AF	BRN3AG	CDKN2O	DDIT13	EPAS18	FANCD13	BRN16
AF25	BRN3AH	BRN3AI	CDKN2P	DDIT14	EPAS19	FANCD14	BRN17
AF26	BRN3AJ	BRN3AK	CDKN2Q	DDIT15	EPAS20	FANCD15	BRN18
AF27	BRN3AL	BRN3AM	CDKN2R	DDIT16	EPAS21	FANCD16	BRN19
AF28	BRN3AN	BRN3AO	CDKN2S	DDIT17	EPAS22	FANCD17	BRN20
AF29	BRN3AP	BRN3AQ	CDKN2T	DDIT18	EPAS23	FANCD18	BRN21
AF30	BRN3AR	BRN3AS	CDKN2U	DDIT19	EPAS24	FANCD19	BRN22
AF31	BRN3AT	BRN3AT	CDKN2V	DDIT20	EPAS25	FANCD20	BRN23
AF32	BRN3AU	BRN3AU	CDKN2W	DDIT21	EPAS26	FANCD21	BRN24
AF33	BRN3AV	BRN3AV	CDKN2X	DDIT22	EPAS27	FANCD22	BRN25
AF34	BRN3AW	BRN3AW	CDKN2Y	DDIT23	EPAS28	FANCD23	BRN26
AF35	BRN3AX	BRN3AX	CDKN2Z	DDIT24	EPAS29	FANCD24	BRN27
AF36	BRN3AY	BRN3AY	CDKN2AA	DDIT25	EPAS30	FANCD25	BRN28
AF37	BRN3AZ	BRN3AZ	CDKN2AB	DDIT26	EPAS31	FANCD26	BRN29
AF38	BRN3BA	BRN3BA	CDKN2AC	DDIT27	EPAS32	FANCD27	BRN30
AF39	BRN3BB	BRN3BB	CDKN2AD	DDIT28	EPAS33	FANCD28	BRN31
AF40	BRN3BC	BRN3BC	CDKN2AE	DDIT29	EPAS34	FANCD29	BRN32
AF41	BRN3BD	BRN3BD	CDKN2AF	DDIT30	EPAS35	FANCD30	BRN33
AF42	BRN3BE	BRN3BE	CDKN2AG	DDIT31	EPAS36	FANCD31	BRN34
AF43	BRN3BF	BRN3BF	CDKN2AH	DDIT32	EPAS37	FANCD32	BRN35
AF44	BRN3BG	BRN3BG	CDKN2AI	DDIT33	EPAS38	FANCD33	BRN36
AF45	BRN3BH	BRN3BH	CDKN2AJ	DDIT34	EPAS39	FANCD34	BRN37
AF46	BRN3BI	BRN3BI	CDKN2AK	DDIT35	EPAS40	FANCD35	BRN38
AF47	BRN3BJ	BRN3BJ	CDKN2AL	DDIT36	EPAS41	FANCD36	BRN39
AF48	BRN3BK	BRN3BK	CDKN2AM	DDIT37	EPAS42	FANCD37	BRN40
AF49	BRN3BL	BRN3BL	CDKN2AN	DDIT38	EPAS43	FANCD38	BRN41
AF50	BRN3BM	BRN3BM	CDKN2AO	DDIT39	EPAS44	FANCD39	BRN42
AF51	BRN3BN	BRN3BN	CDKN2AP	DDIT40	EPAS45	FANCD40	BRN43
AF52	BRN3BO	BRN3BO	CDKN2AQ	DDIT41	EPAS46	FANCD41	BRN44
AF53	BRN3BP	BRN3BP	CDKN2AR	DDIT42	EPAS47	FANCD42	BRN45
AF54	BRN3BQ	BRN3BQ	CDKN2AS	DDIT43	EPAS48	FANCD43	BRN46
AF55	BRN3BR	BRN3BR	CDKN2AT	DDIT44	EPAS49	FANCD44	BRN47
AF56	BRN3BS	BRN3BS	CDKN2AU	DDIT45	EPAS50	FANCD45	BRN48
AF57	BRN3BT	BRN3BT	CDKN2AV	DDIT46	EPAS51	FANCD46	BRN49
AF58	BRN3BU	BRN3BU	CDKN2AW	DDIT47	EPAS52	FANCD47	BRN50
AF59	BRN3BV	BRN3BV	CDKN2AX	DDIT48	EPAS53	FANCD48	BRN51
AF60	BRN3BW	BRN3BW	CDKN2AY	DDIT49	EPAS54	FANCD49	BRN52
AF61	BRN3BX	BRN3BX	CDKN2AZ	DDIT50	EPAS55	FANCD50	BRN53
AF62	BRN3BY	BRN3BY	CDKN2BA	DDIT51	EPAS56	FANCD51	BRN54
AF63	BRN3BZ	BRN3BZ	CDKN2BB	DDIT52	EPAS57	FANCD52	BRN55
AF64	BRN3CA	BRN3CA	CDKN2BC	DDIT53	EPAS58	FANCD53	BRN56
AF65	BRN3CB	BRN3CB	CDKN2BD	DDIT54	EPAS59	FANCD54	BRN57
AF66	BRN3CC	BRN3CC	CDKN2BE	DDIT55	EPAS60	FANCD55	BRN58
AF67	BRN3CD	BRN3CD	CDKN2BF	DDIT56	EPAS61	FANCD56	BRN59
AF68	BRN3CE	BRN3CE	CDKN2BG	DDIT57	EPAS62	FANCD57	BRN60
AF69	BRN3CF	BRN3CF	CDKN2BH	DDIT58	EPAS63	FANCD58	BRN61
AF70	BRN3CG	BRN3CG	CDKN2BI	DDIT59	EPAS64	FANCD59	BRN62
AF71	BRN3CH	BRN3CH	CDKN2BJ	DDIT60	EPAS65	FANCD60	BRN63
AF72	BRN3CI	BRN3CI	CDKN2BK	DDIT61	EPAS66	FANCD61	BRN64
AF73	BRN3CJ	BRN3CJ	CDKN2BL	DDIT62	EPAS67	FANCD62	BRN65
AF74	BRN3CK	BRN3CK	CDKN2BM	DDIT63	EPAS68	FANCD63	BRN66
AF75	BRN3CL	BRN3CL	CDKN2BN	DDIT64	EPAS69	FANCD64	BRN67
AF76	BRN3CM	BRN3CM	CDKN2BO	DDIT65	EPAS70	FANCD65	BRN68
AF77	BRN3CN	BRN3CN	CDKN2BP	DDIT66	EPAS71	FANCD66	BRN69
AF78	BRN3CO	BRN3CO	CDKN2BQ	DDIT67	EPAS72	FANCD67	BRN70
AF79	BRN3CP	BRN3CP	CDKN2BR	DDIT68	EPAS73	FANCD68	BRN71
AF80	BRN3CQ	BRN3CQ	CDKN2BS	DDIT69	EPAS74	FANCD69	BRN72
AF81	BRN3CR	BRN3CR	CDKN2BT	DDIT70	EPAS75	FANCD70	BRN73
AF82	BRN3CS	BRN3CS	CDKN2BU	DDIT71	EPAS76	FANCD71	BRN74
AF83	BRN3CT	BRN3CT	CDKN2BV	DDIT72	EPAS77	FANCD72	BRN75
AF84	BRN3CU	BRN3CU	CDKN2BW	DDIT73	EPAS78	FANCD73	BRN76
AF85	BRN3CV	BRN3CV	CDKN2BX	DDIT74	EPAS79	FANCD74	BRN77
AF86	BRN3CW	BRN3CW	CDKN2BY	DDIT75	EPAS80	FANCD75	BRN78
AF87	BRN3CX	BRN3CX	CDKN2BZ	DDIT76	EPAS81	FANCD76	BRN79
AF88	BRN3CY	BRN3CY	CDKN2CA	DDIT77	EPAS82	FANCD77	BRN80
AF89	BRN3CZ	BRN3CZ	CDKN2CB	DDIT78	EPAS83	FANCD78	BRN81
AF90	BRN3DA	BRN3DA	CDKN2CC	DDIT79	EPAS84	FANCD79	BRN82
AF91	BRN3DB	BRN3DB	CDKN2CD	DDIT80	EPAS85	FANCD80	BRN83
AF92	BRN3DC	BRN3DC	CDKN2CE	DDIT81	EPAS86	FANCD81	BRN84
AF93	BRN3DD	BRN3DD	CDKN2CF	DDIT82	EPAS87	FANCD82	BRN85
AF94	BRN3DE	BRN3DE	CDKN2CG	DDIT83	EPAS88	FANCD83	BRN86
AF95	BRN3DE	BRN3DE	CDKN2CH	DDIT84	EPAS89	FANCD84	BRN87
AF96	BRN3DF	BRN3DF	CDKN2CI	DDIT85	EPAS90	FANCD85	BRN88
AF97	BRN3DG	BRN3DG	CDKN2CJ	DDIT86	EPAS91	FANCD86	BRN89
AF98	BRN3DH	BRN3DH	CDKN2CK	DDIT87	EPAS92	FANCD87	BRN90
AF99	BRN3DI	BRN3DI	CDKN2CL	DDIT88	EPAS93	FANCD88	BRN91
AF100	BRN3DJ	BRN3DJ	CDKN2CM	DDIT89	EPAS94	FANCD89	BRN92
AF101	BRN3DK	BRN3DK	CDKN2CN	DDIT90	EPAS95	FANCD90	BRN93
AF102	BRN3DL	BRN3DL	CDKN2CO	DDIT91	EPAS96	FANCD91	BRN94
AF103	BRN3DM	BRN3DM	CDKN2CP	DDIT92	EPAS97	FANCD92	BRN95
AF104	BRN3DN	BRN3DN	CDKN2CQ	DDIT93	EPAS98	FANCD93	BRN96
AF105	BRN3DO	BRN3DO	CDKN2CR	DDIT94	EPAS99	FANCD94	BRN97
AF106	BRN3DP	BRN3DP	CDKN2CS	DDIT95	EPAS100	FANCD95	BRN98
AF107	BRN3DQ	BRN3DQ	CDKN2CT	DDIT96	EPAS101	FANCD96	BRN99
AF108	BRN3DR	BRN3DR	CDKN2CU	DDIT97	EPAS102	FANCD97	BRN100
AF109	BRN3DS	BRN3DS	CDKN2CV	DDIT98	EPAS103	FANCD98	BRN101
AF110	BRN3DT	BRN3DT	CDKN2CW	DDIT99	EPAS104	FANCD99	BRN102
AF111	BRN3DU	BRN3DU	CDKN2CX	DDIT100	EPAS105	FANCD100	BRN103
AF112	BRN3DV	BRN3DV	CDKN2CY	DDIT101	EPAS106	FANCD101	BRN104
AF113	BRN3DW	BRN3DW	CDKN2CZ	DDIT102	EPAS107	FANCD102	BRN105
AF114	BRN3DX	BRN3DX	CDKN2DA	DDIT103	EPAS108	FANCD103	BRN106
AF115	BRN3DY	BRN3DY	CDKN2DB	DDIT104	EPAS109	FANCD104	BRN107
AF116	BRN3DZ	BRN3DZ	CDKN2DC	DDIT105	EPAS110	FANCD105	BRN108
AF117	BRN3EA	BRN3EA	CDKN2DD	DDIT106	EPAS111	FANCD106	BRN109
AF118	BRN3EB	BRN3EB	CDKN2DE	DDIT107	EPAS112	FANCD107	BRN110
AF119	BRN3EC	BRN3EC	CDKN2DF	DDIT108	EPAS113	FANCD108	BRN111
AF120	BRN3ED	BRN3ED	CDKN2DG	DDIT109	EPAS114	FANCD109	BRN112
AF121	BRN3EE	BRN3EE	CDKN2DH	DDIT110	EPAS115	FANCD110	BRN113
AF122	BRN3EF	BRN3EF	CDKN2DI	DDIT111	EPAS116	FANCD111	BRN114
AF123	BRN3EG	BRN3EG	CDKN2DJ	DDIT112	EPAS117	FANCD112	BRN115
AF124	BRN3EH	BRN3EH	CDKN2DK	DDIT113	EPAS118	FANCD113	BRN116
AF125	BRN3EI	BRN3EI	CDKN2DL	DDIT114	EPAS119	FANCD114	BRN117
AF126	BRN3EJ	BRN3EJ	CDKN2DM	DDIT115	EPAS120	FANCD115	BRN118
AF127	BRN3EK	BRN3EK	CDKN2DN	DDIT116	EPAS121	FANCD116	BRN119
AF128	BRN3EL	BRN3EL	CDKN2DO	DDIT117	EPAS122	FANCD117	BRN120
AF129	BRN3EM	BRN3EM	CDKN2DP	DDIT118	EPAS123	FANCD118	BRN121
AF130	BRN3EN	BRN3EN	CDKN2DQ	DDIT119	EPAS124	FANCD119	BRN122
AF131	BRN3EO	BRN3EO	CDKN2DR	DDIT120	EPAS125	FANCD120	BRN123
AF132	BRN3EP	BRN3EP	CDKN2DS	DDIT121	EPAS126	FANCD121	BRN124
AF133	BRN3EQ	BRN3EQ	CDKN2DT	DDIT122	EPAS127	FANCD122	BRN125
AF134	BRN3ER	BRN3ER	CDKN2DU	DDIT123	EPAS128	FANCD123	BRN126
AF135	BRN3ES	BRN3ES	CDKN2DV	DDIT124	EPAS129	FANCD124	BRN127
AF136	BRN3ET	BRN3ET	CDKN2DW	DDIT125	EPAS130	FANCD125	BRN128
AF137	BRN3EU	BRN3EU	CDKN2DX	DDIT126	EPAS131	FANCD126	BRN129
AF138	BRN3EV	BRN3EV	CDKN2DY	DDIT127	EPAS132	FANCD127	BRN130
AF139	BRN3EW	BRN3EW	CDKN2DZ	DDIT128	EPAS133	FANCD128	BRN131
AF140	BRN3EX	BRN3EX	CDKN2EA	DDIT129	EPAS134	FANCD129	BRN132
AF141	BRN3EY	BRN3EY	CDKN2EB	DDIT130	EPAS135	FANCD130	BRN133
AF142	BRN3EZ	BRN3EZ	CDKN2EC	DDIT131	EPAS136	FANCD131	BRN134
AF143	BRN3FA	BRN3FA	CDKN2ED	DDIT132	EPAS137	FANCD132	BRN135
AF144	BRN3FB	BRN3FB	CDKN2EE	DDIT133	EPAS138	FANCD133	BRN136
AF145	BRN3FC	BRN3FC	CDKN2EF	DDIT134	EPAS139	FANCD134	BRN137
AF146	BRN3FD	BRN3FD	CDKN2EG	DDIT135	EPAS140	FANCD135	BRN138
AF147	BRN3FE	BRN3FE	CDKN2EH	DDIT136	EPAS141	FANCD136	BRN139
AF148	BRN3FF	BRN3FF	CDKN2EI	DDIT137	EPAS142	FANCD137	BRN140
AF149	BRN3FG	BRN3FG	CDKN2EJ	DDIT138	EPAS143	FANCD138	BRN141
AF150	BRN3FH	BRN3FH	CDKN2EK	DDIT139	EPAS144	FANCD139	BRN142
AF151	BRN3FI	BRN3FI	CDKN2EL	DDIT140	EPAS145	FANCD140	BRN143
AF152	BRN3FJ	BRN3FJ	CDKN2EM	DDIT141	EPAS146	FANCD141	BRN144
AF153	BRN3FK	BRN3FK	CDKN2EN	DDIT142	EPAS147	FANCD142	BRN145
AF154	BRN3FL	BRN3FL	CDKN2EO	DDIT143	EPAS148	FANCD143	BRN146
AF155	BRN3FM	BRN3FM	CDKN2EP	DDIT144	EPAS149	FANCD144	BRN147
AF156	BRN3FN	BRN3FN	CDKN2EQ	DDIT145	EPAS150	FANCD145	BRN148
AF157	BRN3FO	BRN3FO	CDKN2ER	DDIT146	EPAS151	FANCD146	BRN149
AF158	BRN3FP	BRN3FP	CDKN2ES	DDIT147	EPAS152	FANCD147	BRN150
AF159	BRN3FQ	BRN3FQ	CDKN2ET	DDIT148	EPAS153	FANCD148	BRN151
AF160	BRN3FR	BRN3FR	CDKN2EU	DDIT149	EPAS154	FANCD149	

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

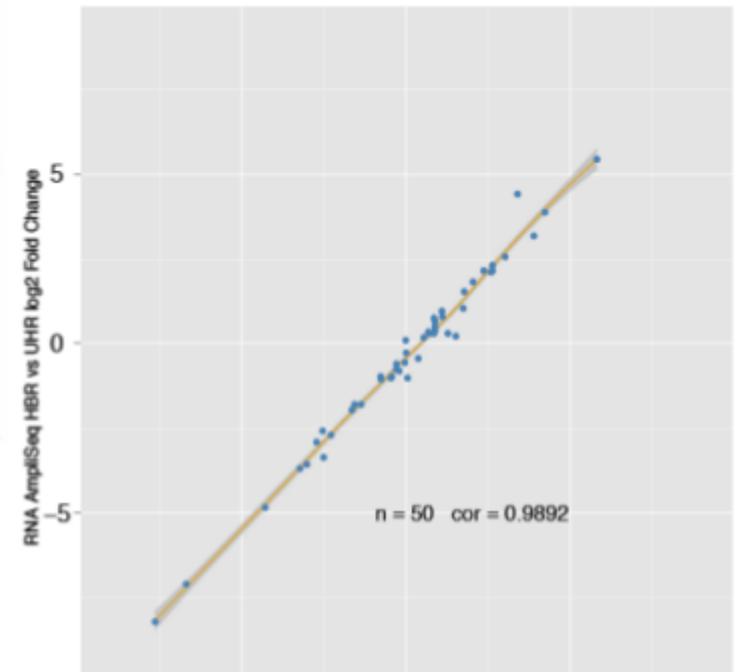
Ion AmpliSeq™ RNA Cancer Panel

50 genes, from 500 pg unfixed RNA, one tube

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

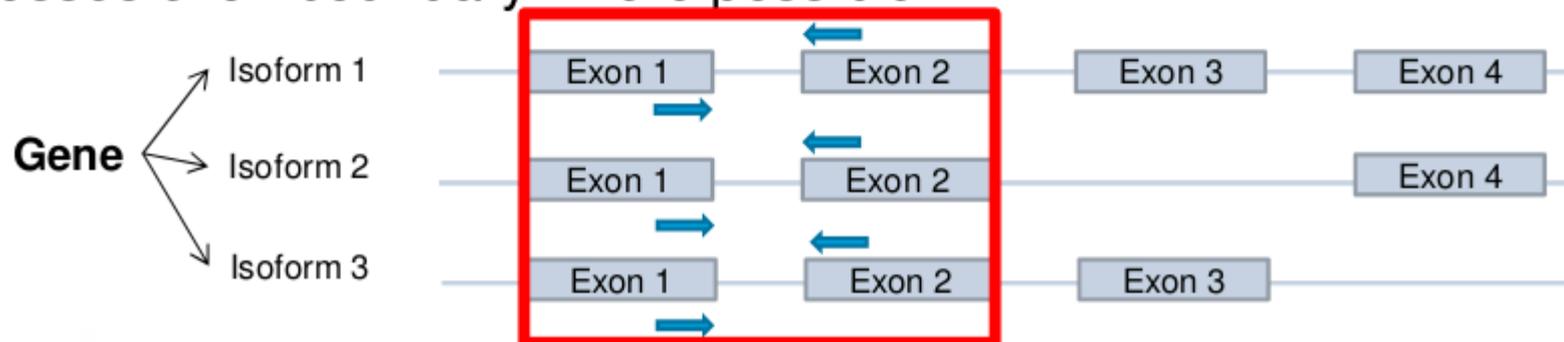
Targeted
quantitative
expression

Complement to
Cancer Hotspot Panel



Ion AmpliSeq™ Transcriptome – Overview

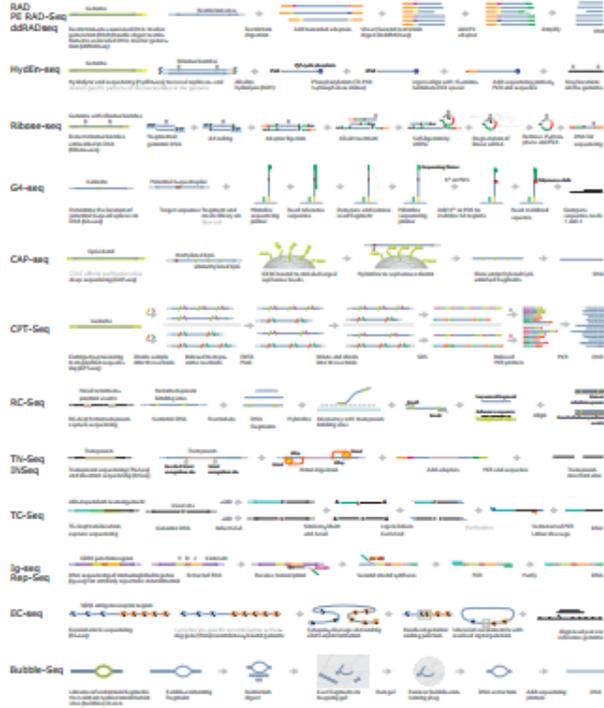
- **Comprehensive** coverage of RefSeq - 20,802 genes targeted
 - Single primer pool – simple workflow
 - Reports at **gene-level**
- **FFPE-compatible** as little as 10 ng RNA
 - Total RNA input from FFPE and other sources – **no selection or enrichment**
- Assay design
 - One amplicon per gene
 - ~150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

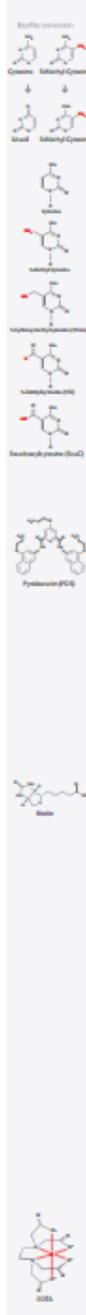
DNA Rearrangements and Markers



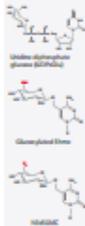
DNA Low-Level Detection



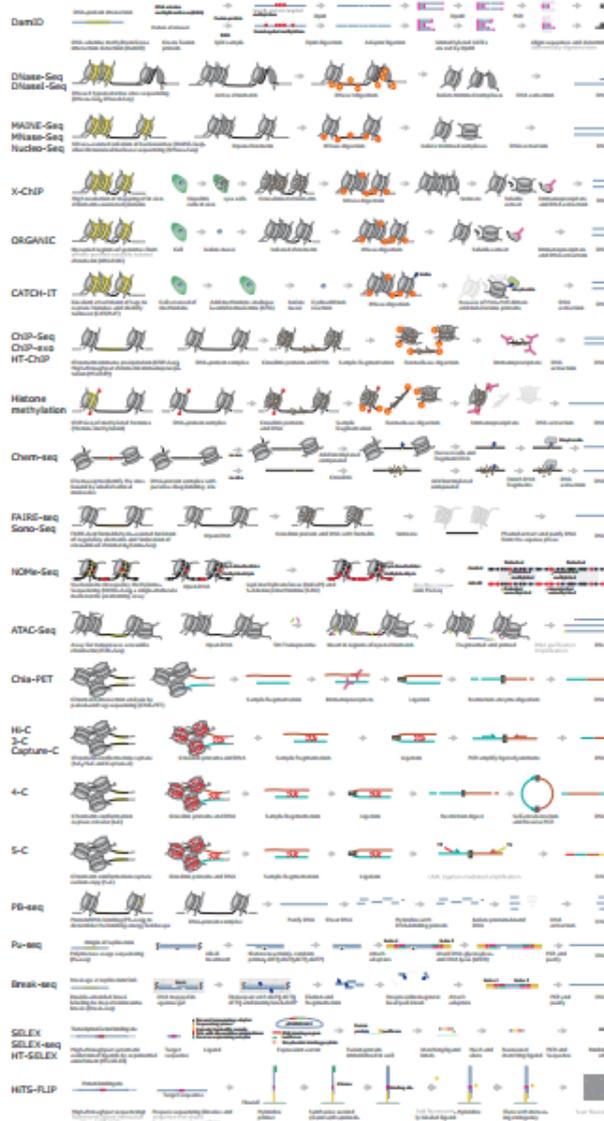
Key



Epigenetics



DNA-Protein Interactions



Protein-Protein Interactions



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.

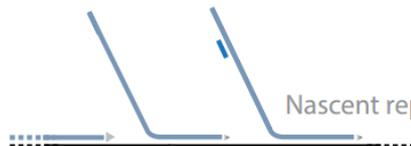
- In this method, 3'-blocked random hexamer primers are hybridized to the template, followed by synthesis with Phi 29 polymerase.
- Phi 29 performs strand-displacement DNA synthesis, allowing for efficient and rapid DNA amplification.

3' blocked random hexamer primers



Hybridize primers

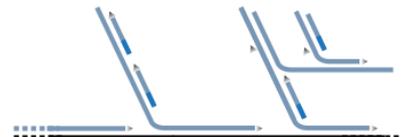
Phi 29



Nascent replication fork

Synthesis

Phi 29



Synthesis

S1 nuclease



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.

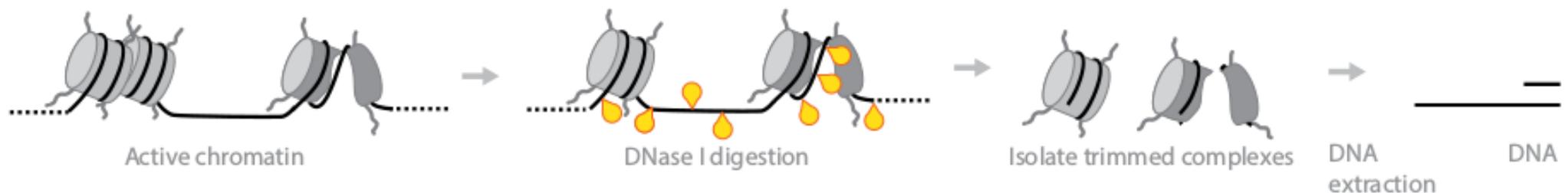
Histone modifications determine gene activation by recruiting regulatory factors and maintaining an open or closed chromatin state. Epigenetic factors play roles in tissue development, embryogenesis, cell fate, immune response, and diseases such as cancer. Bacterial pathogens can elicit transcriptional repression of immune genes by chromatin remodeling.



Cigarette smoking disrupts DNA-protein interactions leading to the development of cancers or pulmonary diseases.

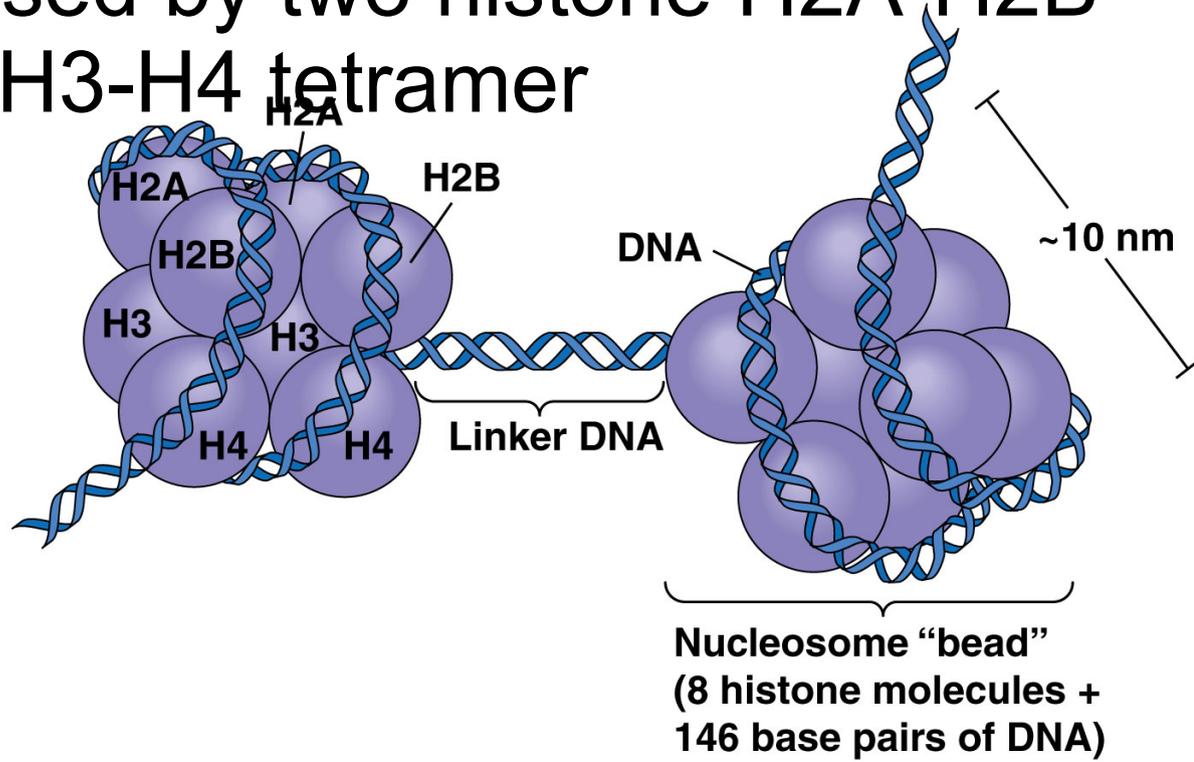
DNASE I HYPERSENSITIVE SITES SEQUENCING (DNASE-SEQ)

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

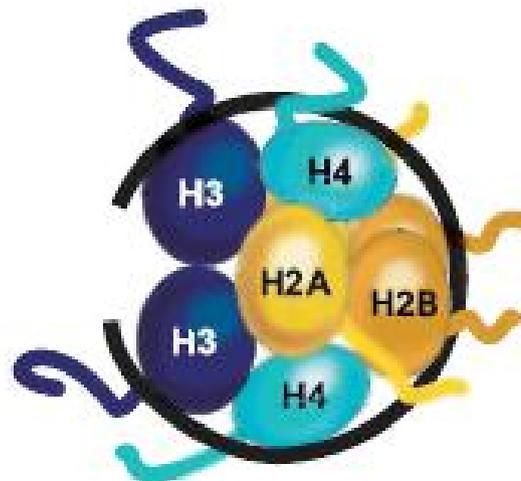
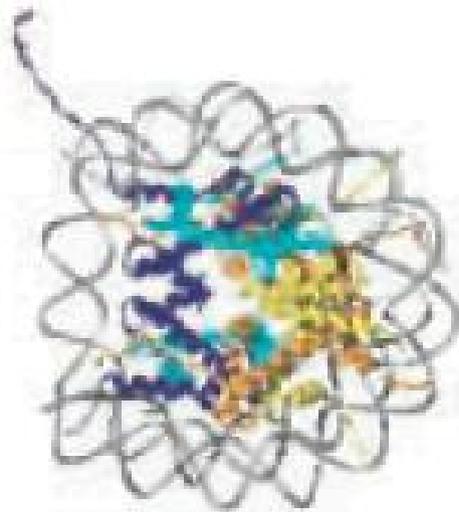


Chip-seq

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

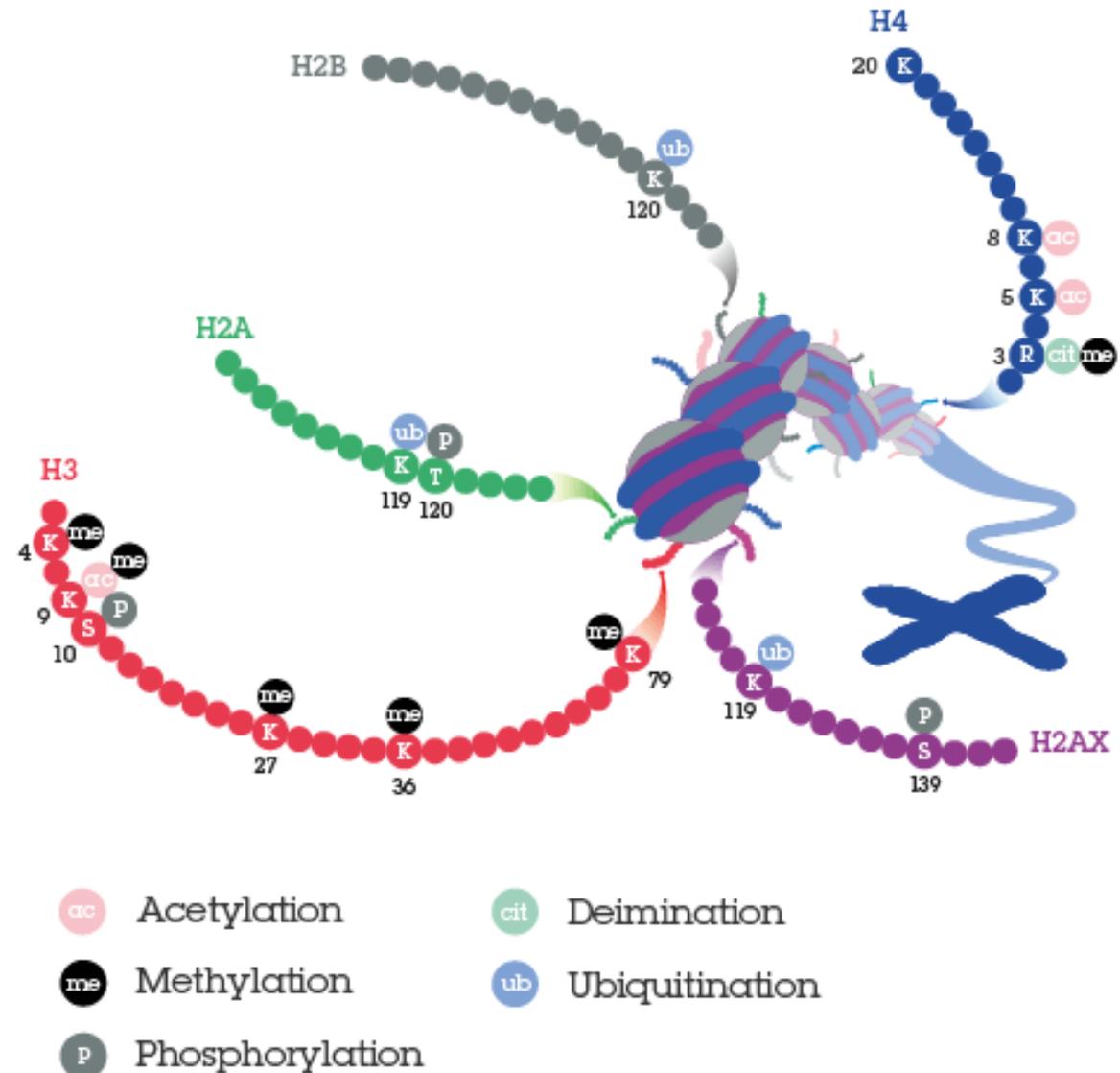


Each histone protein contains the so-called histone fold structural motif, flanked by unstructured N- and C-terminal tails, ranging from 15 (H2A) to 35 (H3) amino acids, that protrude from the nucleosomal core.



Histone modifications

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



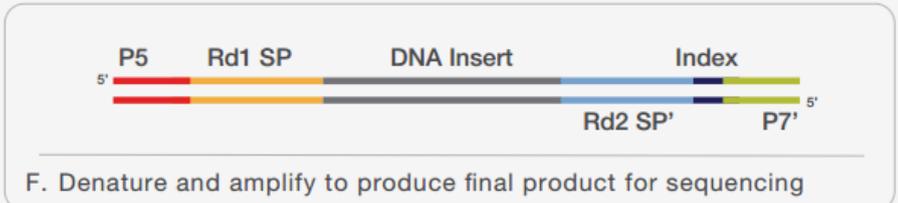
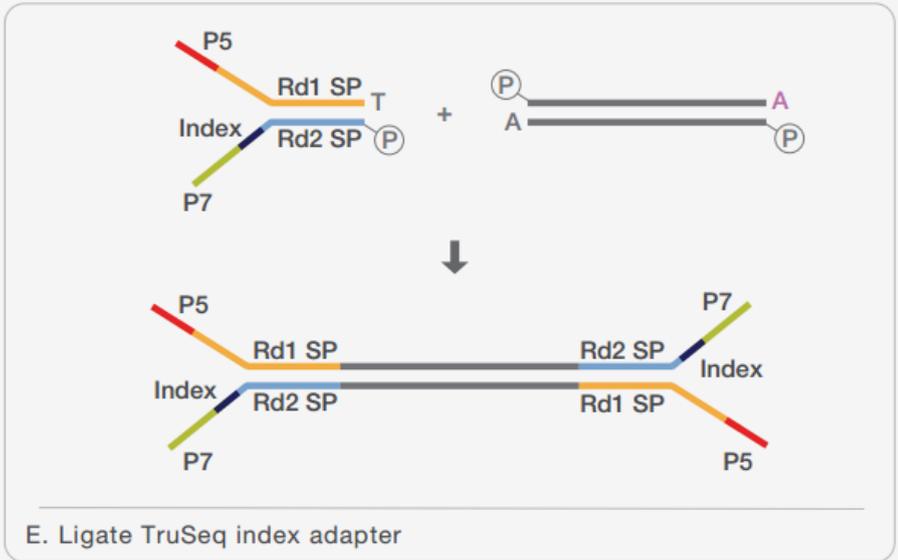
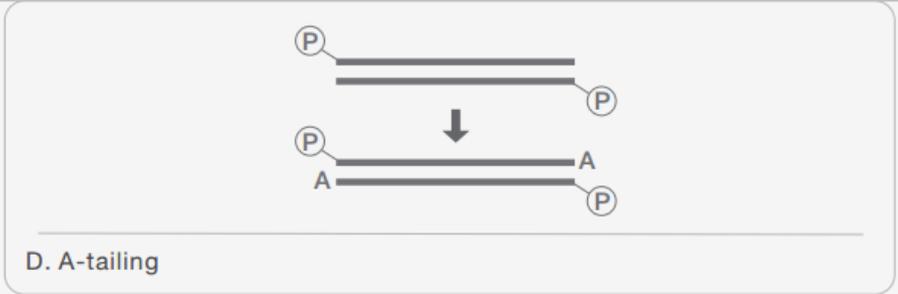
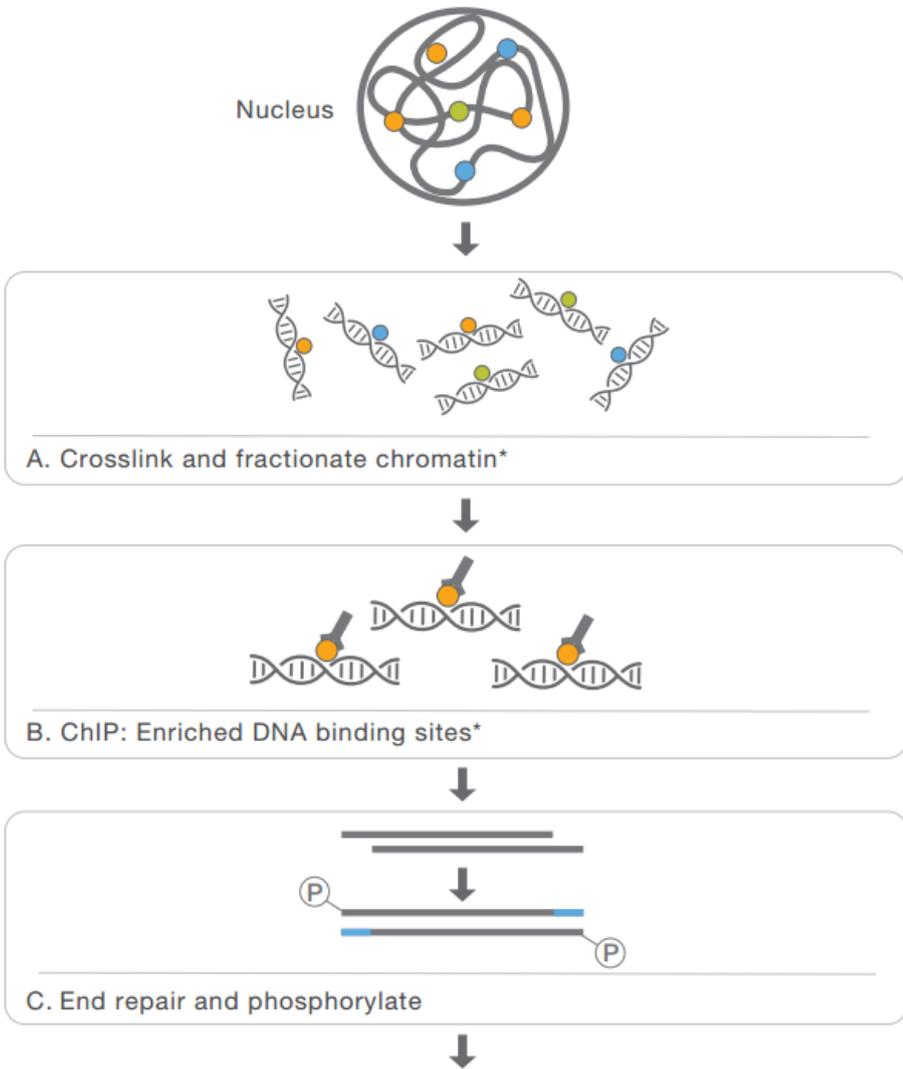
H3K4, H3K36, and H3K79 methylations are generally linked to active gene expression, whereas H3K9, H3K27, and H4K20 di- and tri-methylations have been associated with gene silencing.

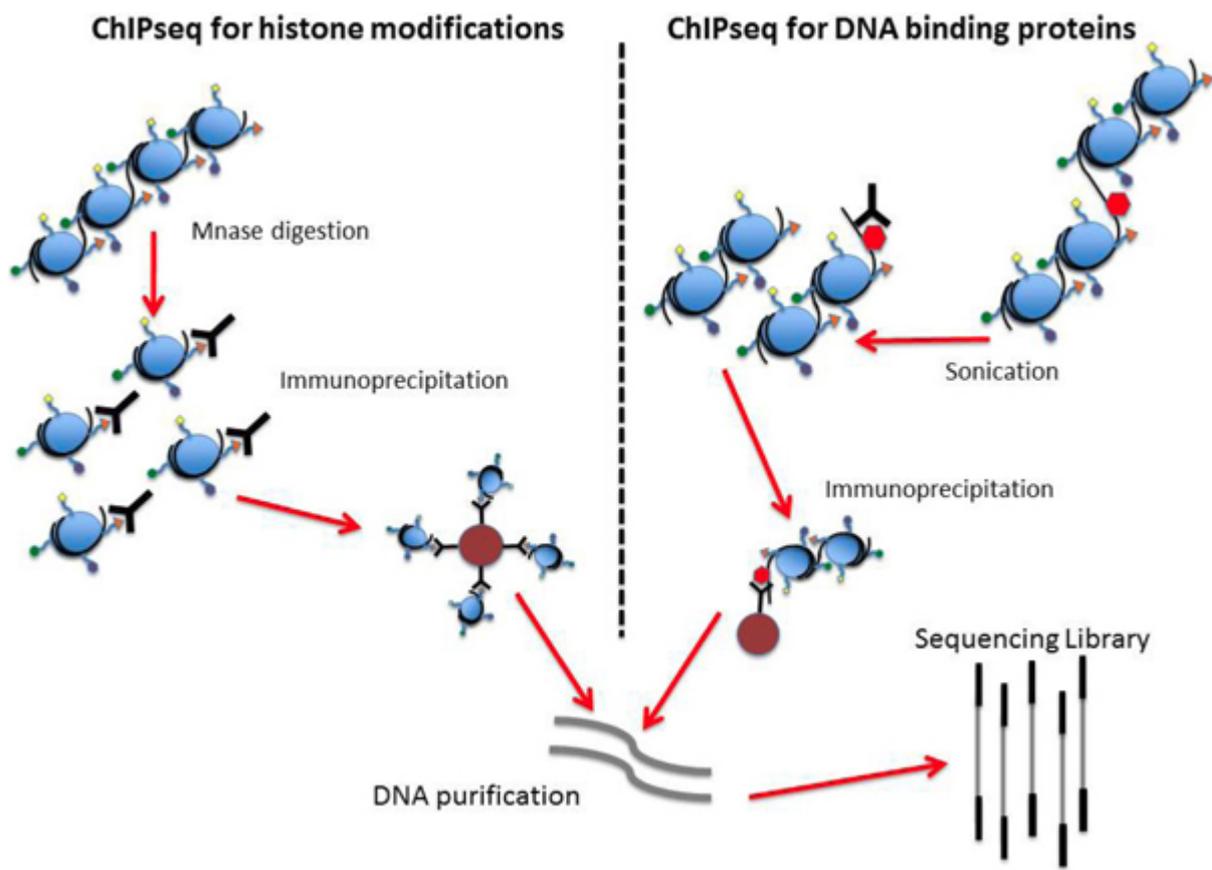
H3K4me2 and H3K4me3 have been both found predominantly on active loci, although H3K4me3 is associated with active genes, H3K4me2 can be present also in inactive genes

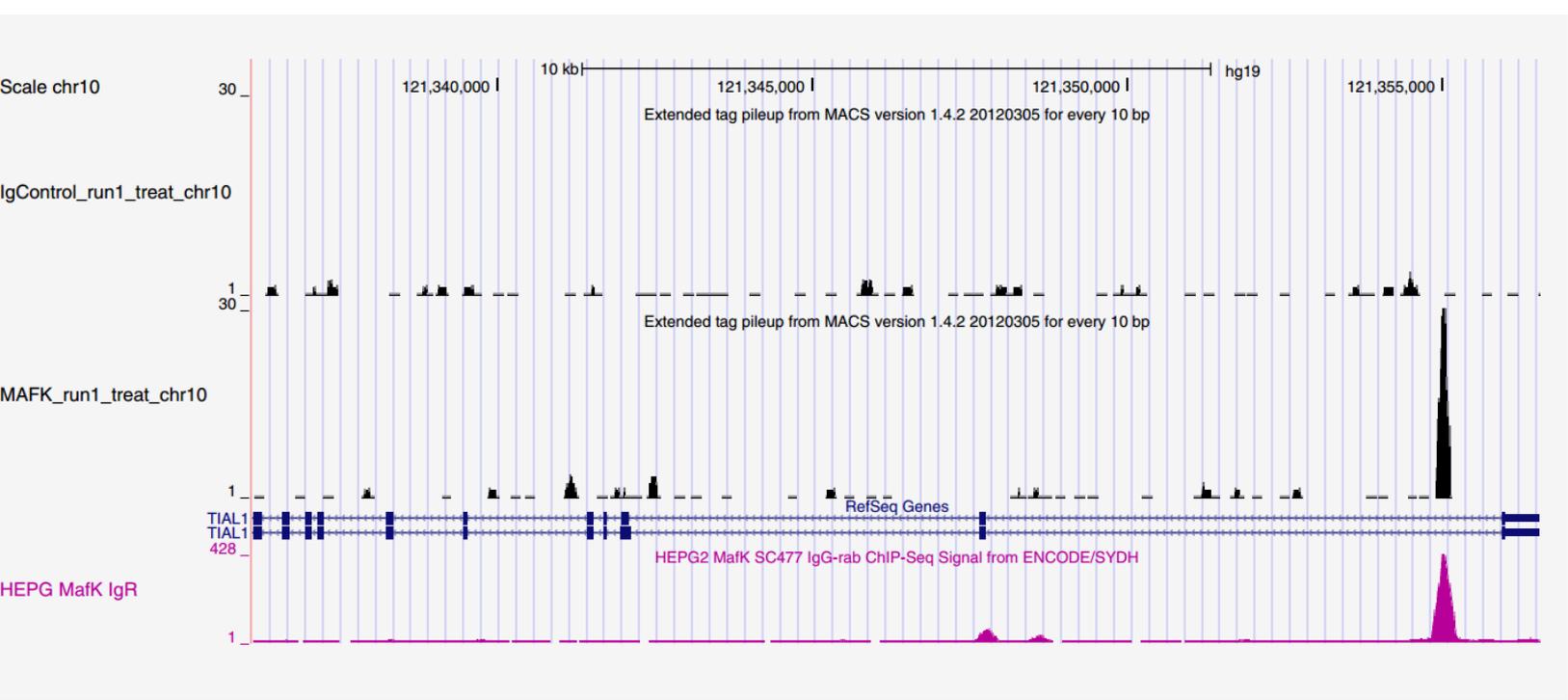
TruSeq® ChIP Sample Preparation

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

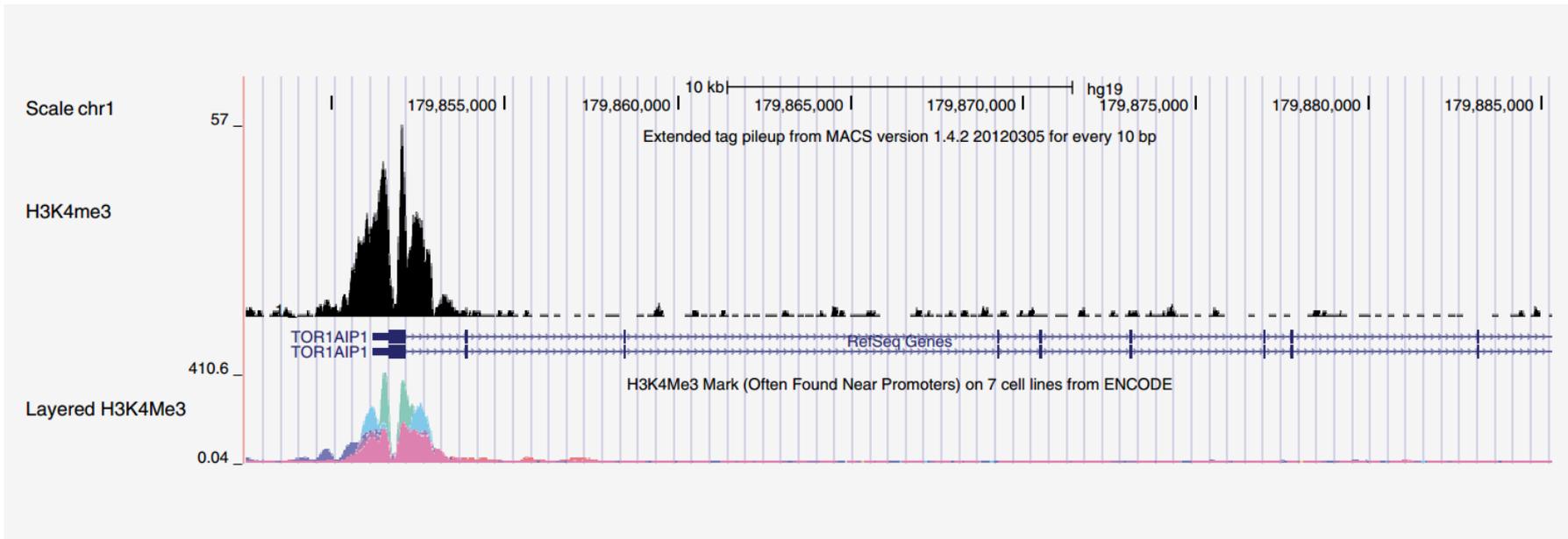






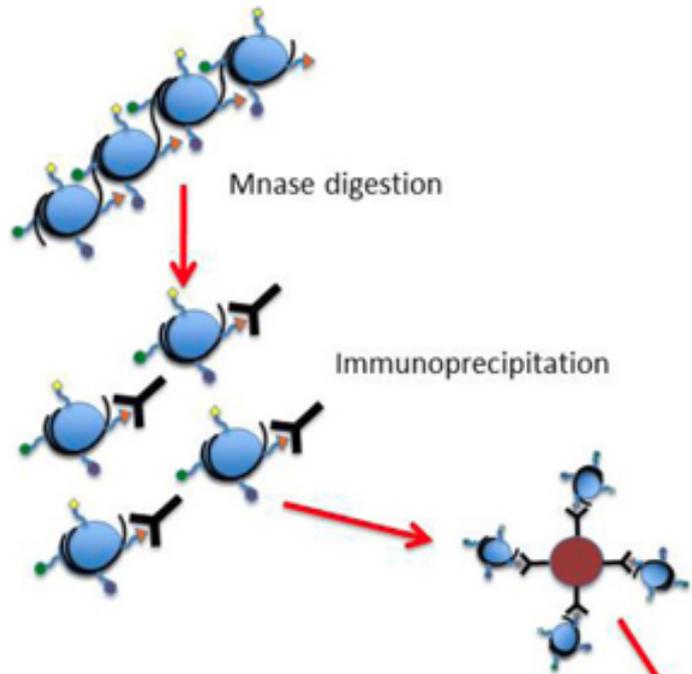
TruSeq ChIP Sample Preparation Kits enable the generation of libraries across a broad range of study designs. Above is peak data for a negative Ig control, the transcription factor target MafK, and a reference peak for MafK from the ENCODE database.

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

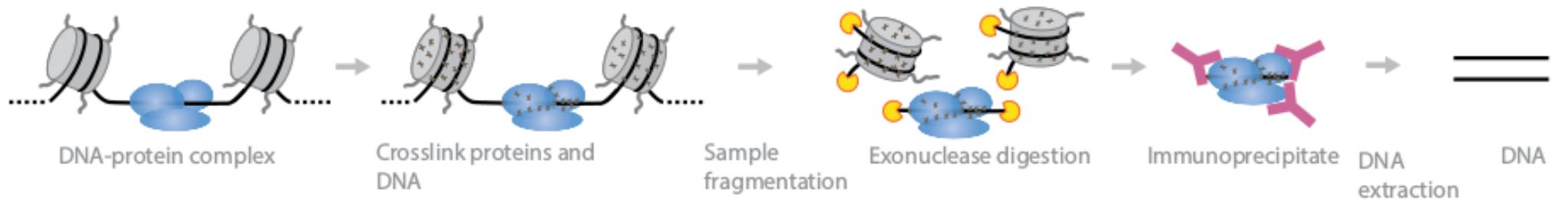
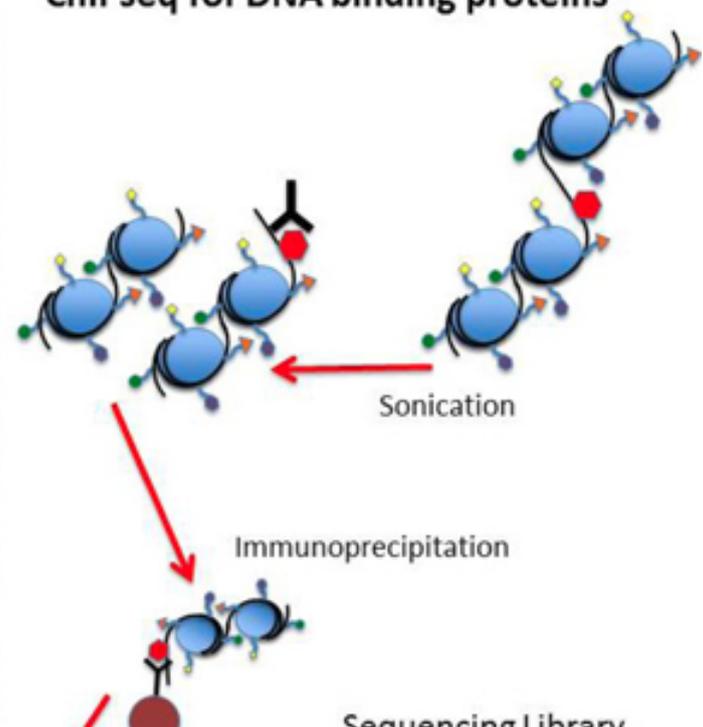


The peak results for the H3K4me3 target compare favorably with the ENCODE annotation data for this well characterized target, with a representative peak for the histone mark target H3K4me3 and a corresponding ENCODE reference peak.

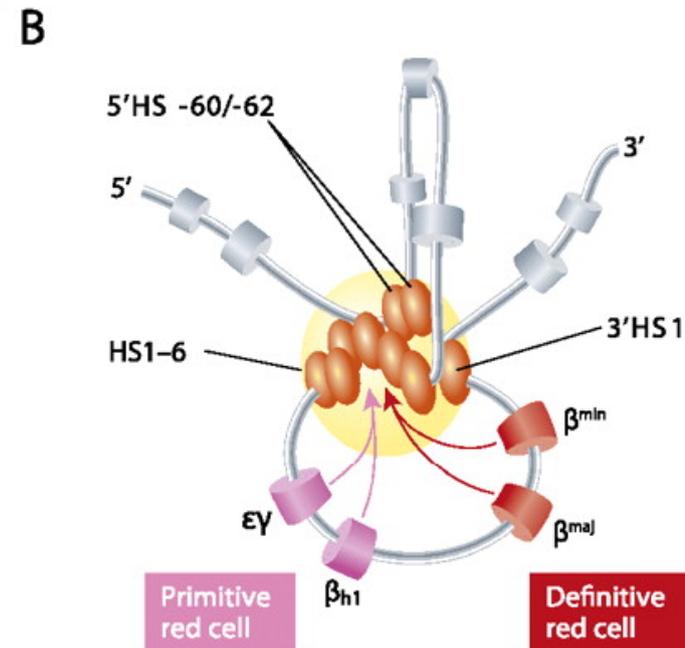
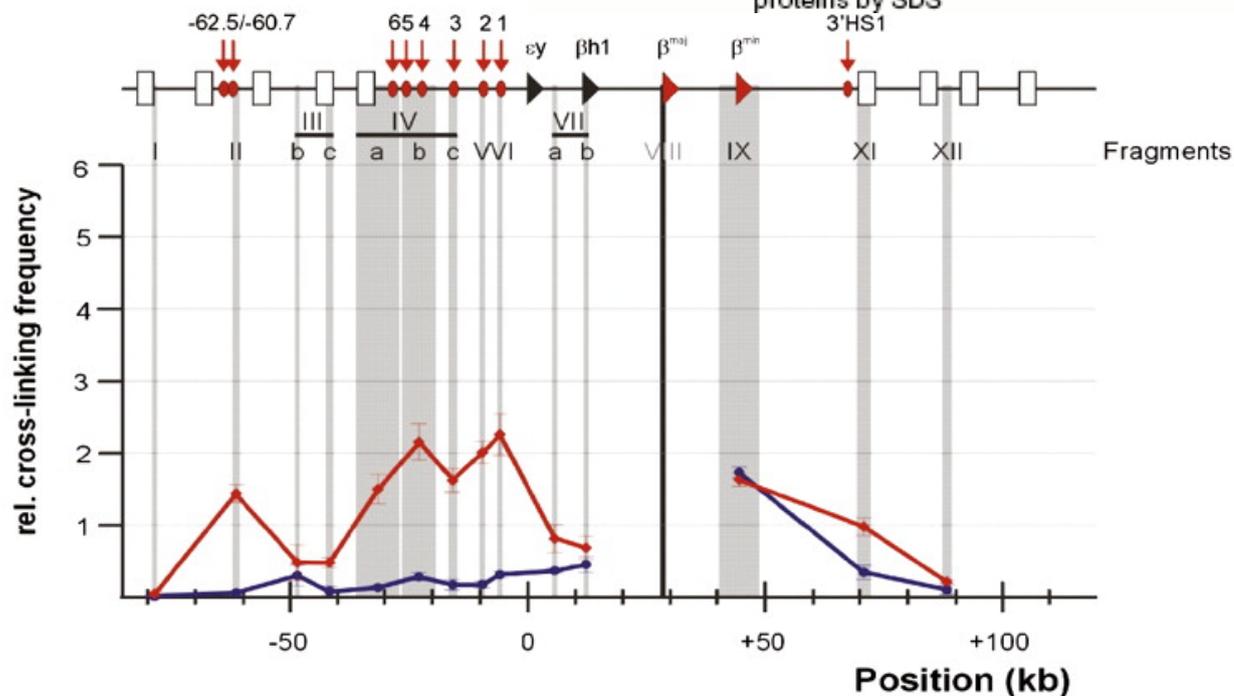
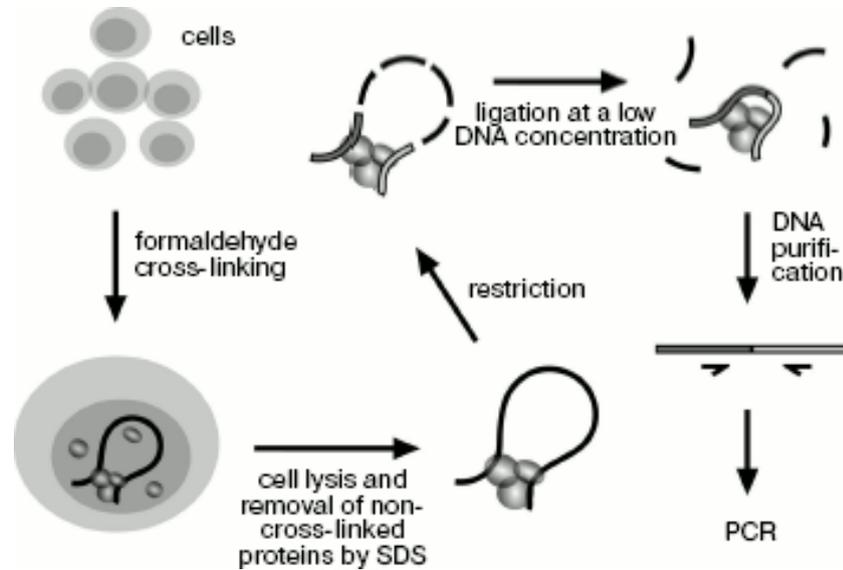
ChIPseq for histone modifications



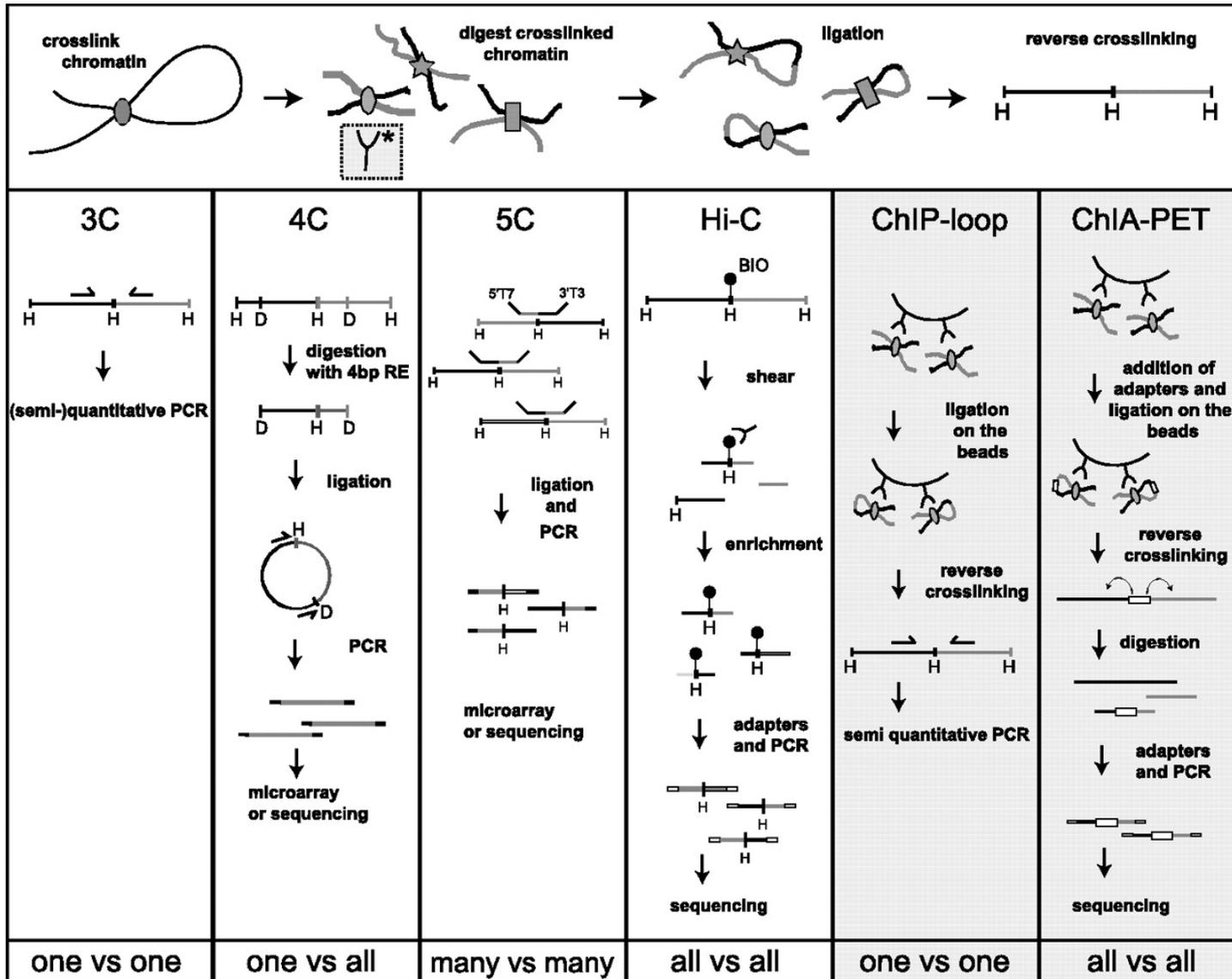
ChIPseq for DNA binding proteins



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

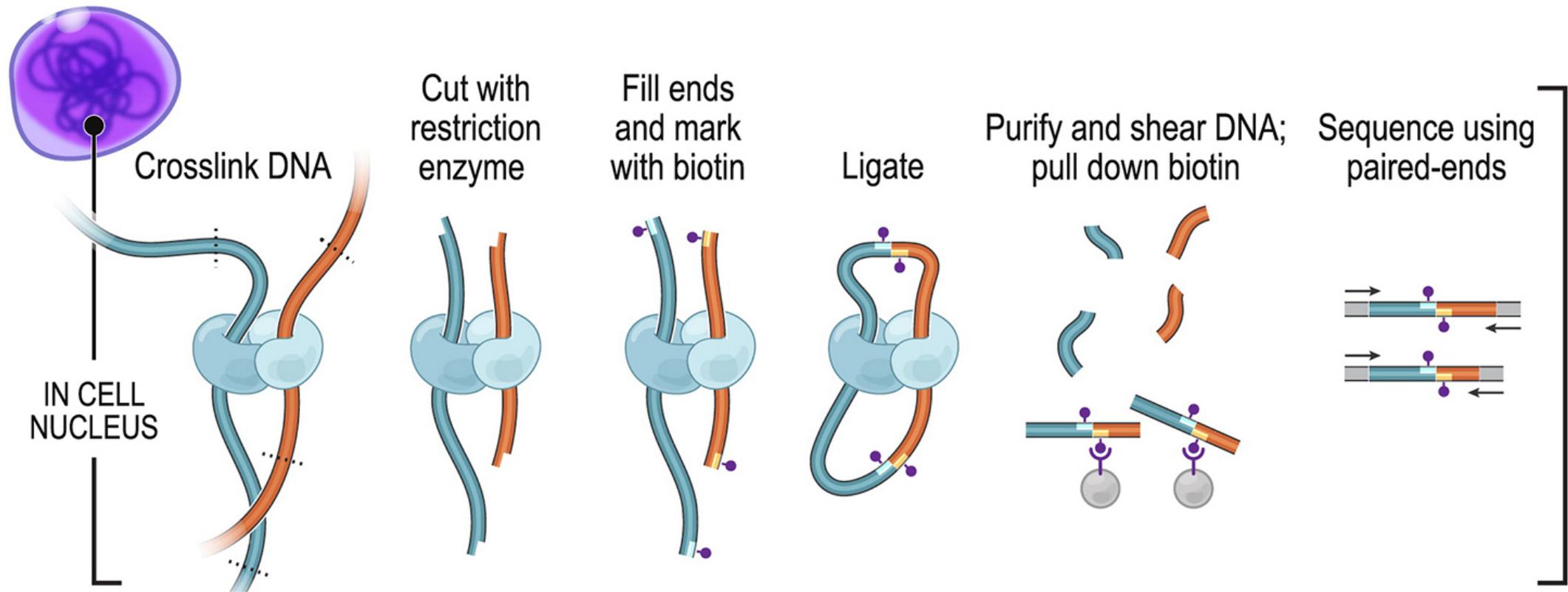


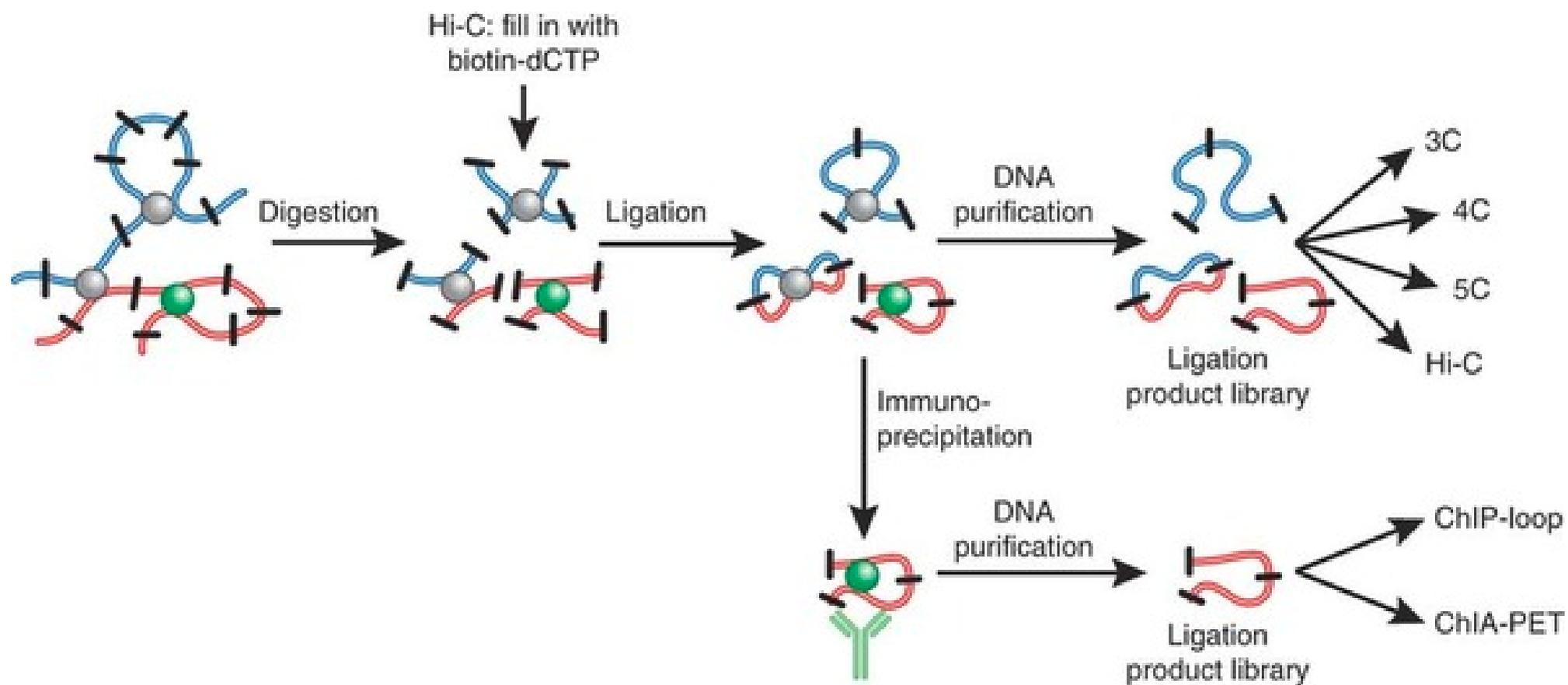
Overview of 3C-derived methods.



Elzo de Wit, and Wouter de Laat *Genes Dev.* 2012;26:11-24



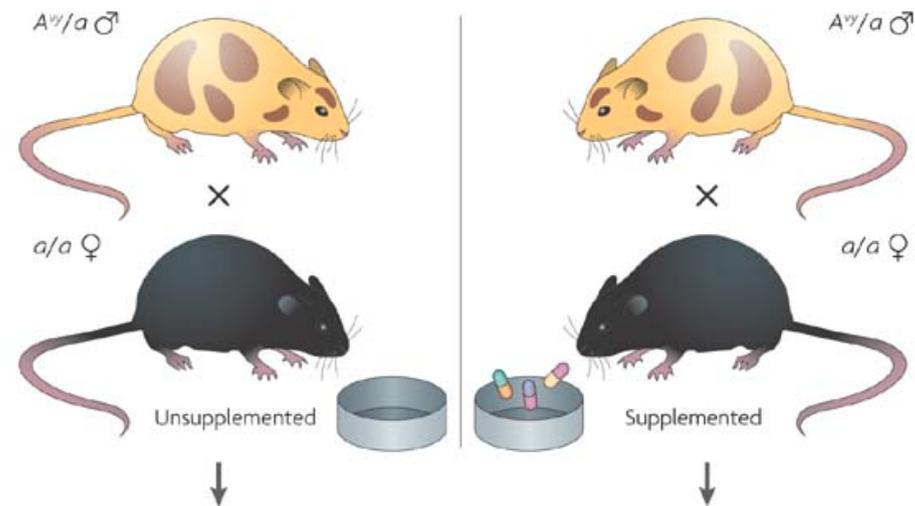




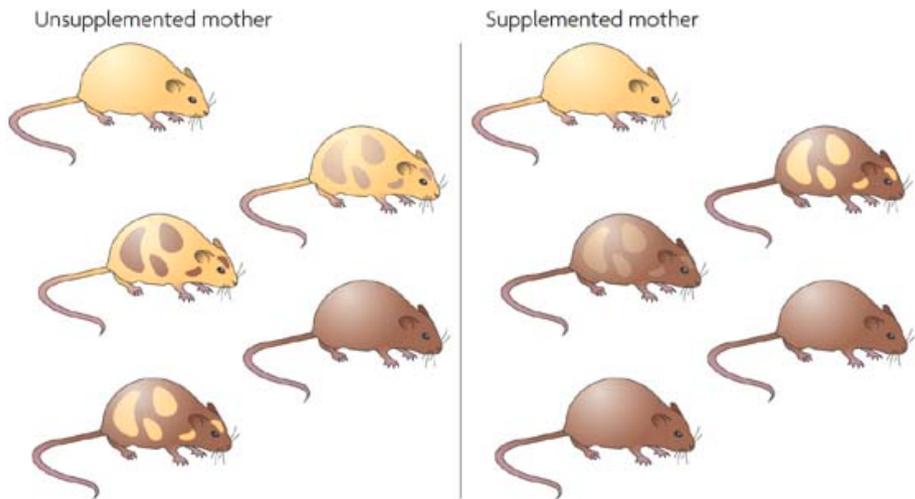
DNA Methylation

- 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.

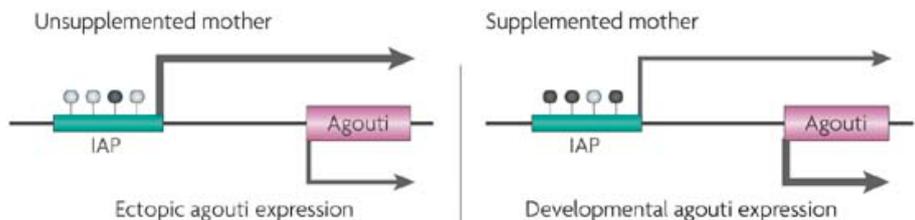
a Dietary supplementation during pregnancy



b A^y/a offspring



c Agouti expression



- 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

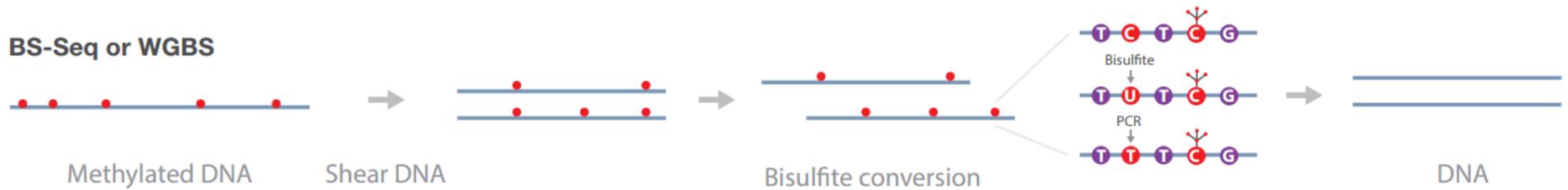
Nature Reviews Genetics 8, 253-262 (April 2007)

doi:10.1038/nrg2045

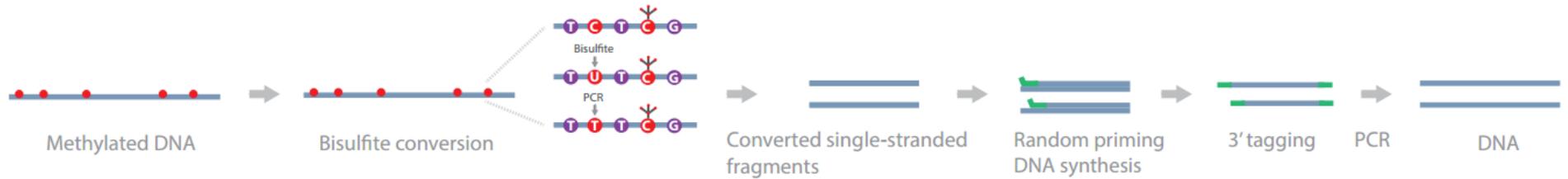
- Most cytosine methylation occurs on cytosines located near guanines, called CpG sites. These CpG sites are often located upstream of promoters, or within the gene body. CpG islands are defined as regions that are greater than 500 bp in length with greater than 55% GC and an expected/observed CpG ratio of > 0.65 .
- While cytosine methylation (5mC) is known as a silencing mark that represses genes, cytosine hydroxymethylation (5hmC) is shown to be an activating mark that promotes gene expression and is a proposed intermediate in the DNA demethylation pathway. Similar to 5mC, 5hmC is involved during development, cancers, cell differentiation, and diseases.
- 5mC and/or 5hmC can be a diagnostic tool to help identify the effects of nutrition, carcinogens, and environmental factors in relation to diseases.
- The impact of these modifications on gene regulation depends on their locations within the genome. It is therefore important to determine the exact position of the modified bases.

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 deamination and are read as cytosines. The location of the methylated cytosines can then be determined by comparing treated and untreated sequences.



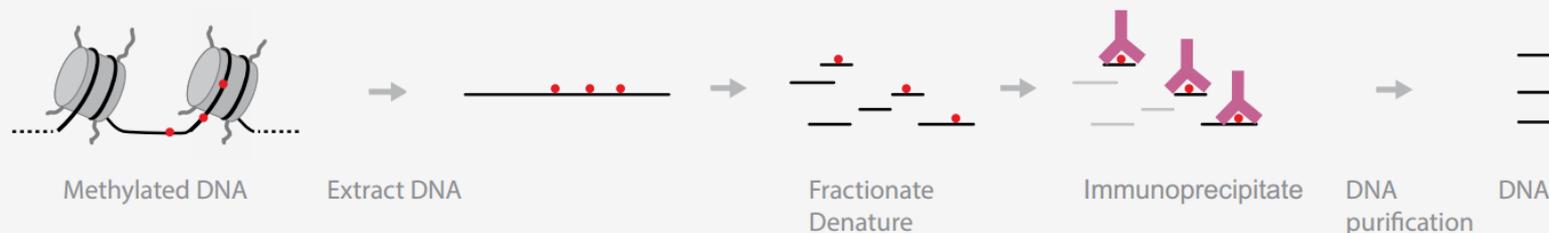
EpiGnome Methyl-Seq



- 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

METHYLATED DNA IMMUNOPRECIPITATION SEQUENCING (MEDIP-SEQ)

- Methylated DNA immunoprecipitation sequencing (MeDIP-Seq) is commonly used to study 5mC or 5hmC modification.
- Specific antibodies can be used to study cytosine modifications. If using 5mC-specific antibodies, methylated DNA is isolated from genomic DNA via immunoprecipitation.
- Anti-5mC antibodies are incubated with fragmented genomic DNA and precipitated, followed by DNA purification and sequencing.



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 arrangement of the sex determining region over evolution, giving us distinct X and Y chromosomes.

