RNASeq Wet-Lab Workflow

Introduction to RNA-seq

What is RNA-seq?

RNA-seq (RNA-sequencing) is a technique that can examine the quantity and seqences of RNA in a sample using next generation sequencing (NGS). It analyzes the transcriptome of gene expression patterns encoded within our RNA.

What are the applications of RNA-seq?

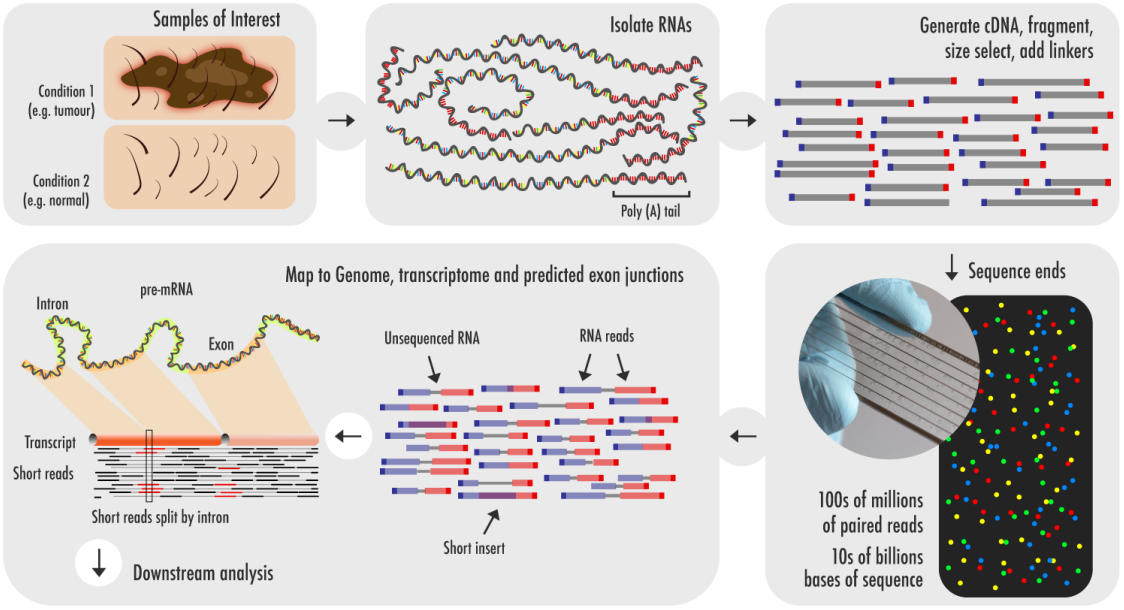
RNA-seq lets us investigate and discover the transcriptome, the total cellular content of RNA's including mRNA, rRNA and tRNA. Understanding the transcriptome is key if we are to connect the informatin on our genome with its functional expression. RNA-seq can tell us which genes are turned on in a cell, what their level of expression is, and what times they are activated or shut off. This allows to more deeply understand the biology of a cell and assess changes that my indicate disease. Some of the most popular techniques that use RNA-seq are transcriptional profiling, SNP identification, RNA editing and differential gene expression analysis.

How does RNA-seq work?

The first step in the technique involves converting the population of RNA to be sequenced into cDNA fragments (a cDNA library). This allows the RNA to be put into an NGS workflow. Adapters are then added to each end of the fragments. These adapters contain functional elements which permit sequencing; for example, the amplification element and the primary sequencing site. The cDNA library is then analyzed by NGS, producing short sequences which correspond to either one or both ends of the fragment. The depth to which the library is sequenced varies depending on techniques which the output data will be used for. The sequencing often follows either single-read or paired-end sequencing methods. Single-read sequencing is cheaper and faster.

A further choice must be made between strand-specific and non-strand-specific protocols. The former method means the information about which DNA strand was transcribed is retained. The value of extra information obtained from strand-specific protocols make them the favorable option.

These reads, of which there will be many millions by the end of the workflow, can then be aligned to a genome of reference and assembled to produce an RNA sequence map that spans the transcriptome.



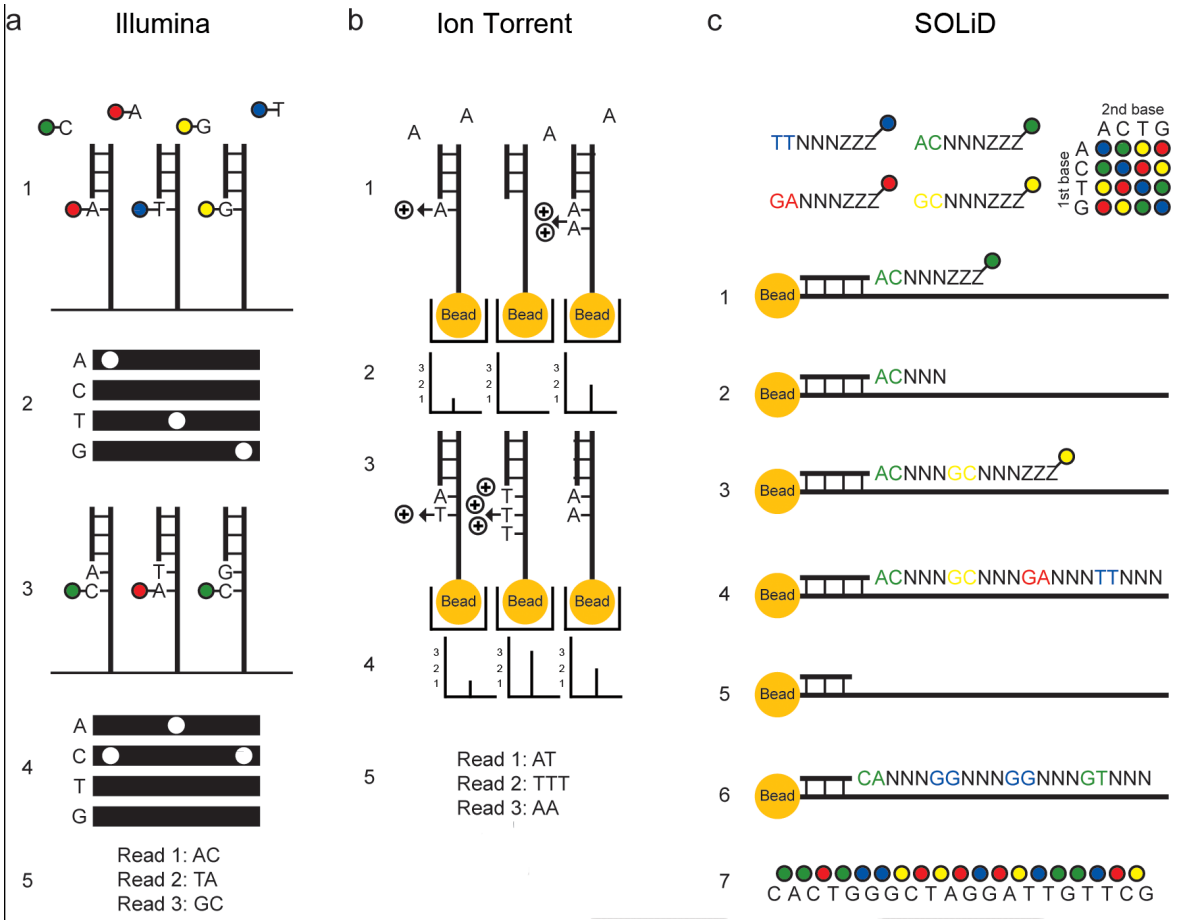
The general Workflow

1. **Sequencing (biochemistry)**
   A. RNA extraction
   B. Library preparation (including mRNA enrichment)
   C. Sequencing

2. **Bioinformatics**
   A. Processing of sequencing reads (including alignment)
   B. Estimation of individual gene expression levels
   C. Normalization
   D. Identification of differentially expressed (DE) genes

RNAseq Technologies

Most sequencing platform support multiplexing of libraries by introducing barcodes during library preparation. This allows simultaneous sequencing of multiple libraries in a single sequencing run thereby enabling more efficient use of a sequencer machine. Importantly, multiplexing also facilitates a balanced block design to minimize potential confounding factors such as PCR amplification and flow cell effects.



Illumina

Illumina and Ion Torrent both sequence using a "sequencing by synthesis" (SBS) approach, whereby incorporation of dNTPs is detected simultanously at millions of fixed positions on a flow cell.

For Illumina, once RNA-seq libraries have been prepared, they are hybridized to a flow cell which contains a lawn of covalently bound oligonucleotides complementary to the sequencing adapters that were introduced during library preparation. Once hybridized, the capture oligonucleotide primes DNA polymerase extension activity resulting in a covalently bound full-length complementary copy of the cDNA fragment that is subjected to several rounds of PCR amplification to produce discrete clones ~1µm in diameter that can be optically resolved during sequencing. Obtaining optimal cluster density is critical since it will determine the number of reads obtained. Clearly, low density will result in fewer than expected reads, but over-clustering can be just as problematic since dense flow cells are difficult to analyze and to obtain accurate base calling due to interference and overlap between adjacent clusters. Therefore, accurate quantification PCR is an important aspect of library quality control.

In the case of Illumina SBS, all four dNTPS are flourescently labeled and concurrently introduced on to the flow cell. Since all four dNTPs are present, natural competition for binding between dNTPs minimizes incorporation biases. SBS proceeds through multiple cycles of nucleotide incorporation and detection. Importantly, only one nucleotide is incorporated per cycle by use of reversibly terminated dNTPs. After nucleotide incorporation is detected by fluorescence, the fluorophore is removed resulting in regeneration of a 3' hydroxyl group which allows incorporation of the next dNTP in the subsequent cycle. Importantly , this reversible terminator chemistry allows sequencing of homoplymeric regions, such as "AAAAAA", with high confidence.

During base calling, fluorescence intensity values for each nucleotide are converted to nucleotide identity using a cross-talk matrix which controls for spectral overlap. Since spectral overlap is determined during the first four cycles, it is imperative that approximately equal numbers of each base be present (i.e., to have a balanced library). Therefore, it is especially important to use barcodes that are well balanced to ensure accurate demultiplexing after sequencing.

Ion Torrent

Whereas Illumina sequencing and cluster generation relies on solid-phase PCR amplification, emulsion PCR is used to prepare Ion Torrent libraries for sequencing. First, the library template is prepared from fragmented RNA. Unlike Illumina, the standard library protocol is strand-specific by default. Next, beads with complementary oligonucleotides are mixed with PCR reagents and a dilute solution of cDNA and oil added to make an emulsion. Ideally, each microdroplet of emulsion will contain one bead and one cDNA fragment along with PCR reagents to allow for clonal amplification. Following 16-18 cycles of PCR the emulsion is then broken by organic extraction, beads purified and loaded on to a disposable semiconductor sequencing chip. The sequencing chip is modeled similar to honeycomb, in that one bead fits into one of hundreds of millions of tiny wells that serve as microreactors during sequencing, each with their own detector. Unlike Illumina's fluorescence-based SBS, Ion Torrent determines sequence identity by detecting pH alterations due to hydrogen ion release following nucleotide incorporation. Since the dNTPs are not differentially labeled by a fluorophore, they must be added successivly so that ion release can be associated with a particular nucleotide. Since Ion Torrent sequencing isn't reliant on optical detection of dNTP incorporation, sequencing reactions are much faster, and the number of reads obtainable per sequencing run has been rapidly increasing. However, whereas Illumina makes use of reversible terminator chemistry to restrict dNTP incorporation to once per cycle and sequence through homopolymers, Ion Torrent relies on the number of hydrogen ions released as being proportional to the number of dNTPs incorporated. Therefore, "A" can easily ve distinguished from "AA" by a detectin a doubling in the number of hydrogen ions released. However, distinguishing between a run of 7 and 8 adenosines is for more challenging, and consequently, the error rate is high.

Sequencing depth and coverage

Technically, *coverage* refers to the number of reads being sequenced in realtion to the genome size, i.e., it is an estimate of how many times each base of the genome is sequenced. To identify sequencing errors (and possibly distinguish them from genomic variants), every base should be covered more then once. For RNA-seq, the coverage estimation has rather little practical value as the size of the transcriptome is not known as accurately as the size of the genome, and, more importantly, on their expression. Thus the number of required reads is determined by the least abundant RNA species of interest. However, it is impossible to know before sequencing how many reads are going to be needed to capture enough fragments of the the most lowly expressed genes. Keep in mind that strongly expressed genes and residual rRNA will always account for large fraction of all reads. In most cases of differential gene expression analysis, it is more imporThe coverage value will alwayse be an estimate as the genome is usually not covered uniformily since, for example, tant to increase the number of biological replictates then the sequencing depth of single samples.

20-50 million single-end read seqeuncing depths are recommended for human samples.

Experimental Design

Most RNA-seq experiments aim to identify genes whose expression varies between two or more experimental settings. This means, that during downstream analyses, we will test every single gene whether its expression seems to change when comparing two (or more) conditions. It seems immediately obvious that comparing just one measurment per condition is not going to yield a very robust answer since gene expression may vary because of many factors (e.g., temperature, sex, time of day), not just because to the condition of interest (e.g., genotype or drug treamtment). To distinguish transcription changes caused by the condition being studied from transcription variation caused by differences between individual organisms, cell populations, or experimenters, it is important to perform RNA-seq experiments with sufficient numbers of different types of replicates and with a well thought-out experimental design. The goal is to observe a reproducible effect that can be due only to the treatment (avoiding confounding and bias) while simulataneously measuring the variability required to estimate how much we expect the effect to differ if the measurements are repeated with similar but not identical samples (replicates).

	Replicate type	Category
Subjects	Colonies	Biological
	Strains	Biological
	Cohoused groups	Biological
	Gender	Biological
	Individuals	Biological
Sample preparation	Organs from sacrificed animals	Biological
	Methods for dissociating cells from tissue	Technical
	Dissociation runs from given tissue sample	Technical
	Individual cells	Biological
	RNA-seq library construction	Technical
Sequencing	Runs from the library of a given cell	Technical
	Reads from different transcript molecules	Variable
	Reads with unique molecular identifier from a given transcript molecule	Technical

Without a somewhat realistic estimate of the variance in your system of interest, the statistical tests will have a very hard time to make accurate inferences about the gene expression differences. Ideally, there should be enough replicates to capture the breatdh of the variability and to identify and isolate sources of noise. In practical terms, this usually translates to a number of replicates that allow to a) identify outlier samples and b) be able to remove them without losing too much information about the background variation between transcripts of the same type. The latter step should only be taken if there are valid reasons to believe that a certain sample might indeed be an outlier due to technical reasons (e.g., sequencing problems) or biological reason that do not play a role for the question at hand.

Technical replicates

Every experiment will have some random noise associated with protocols or equipment. Generally speaking, technical replicates are therefore repeated measurements of the same sample. For RNA-seq specifically, the ENCODE consortium has defined technical replicates as *different library preparations from the same RNA sample*. They should account for batch effects from the libray preparation such as reverse transcription and PCR amplification. In most cases, technical variability introduced by the sequencing protocol is quite low and well controlled, so that technical replicates are rarely done.

Biological replicates

There is an on-going debate over what kinds of samples represent true biological replicates, but generally accepted definition is that biological replicates should be "parallel measurements of biologically distinct sample s that capture random biological variation". Biological replicates will allow you to have a better handle on the true mean and variance of expression (of all genes in question) for the biological population of interest. The ENCODE consortium specifies that biological replicates should represent *RNA from and independent growth of cells/tissue*.

Numbers of replicates

Currently, the most published RNA-seq experiments contain three biological replicates. If the goal of the experiment is to identify as many differntially expressed genes as possible (including slightly changing ones and those that are lowly expressed), as many as twelve replicates are recommended. As a genral rule, the more genes with low fold changes that are to be detected, the more replicates are needed to increase the precision of the estimation of the biological variability.

Impact of sequencing depth and number of replicates

Statistical power to detect differential expression varies with effect size, sequencing depth and number of replicates

	Replicates per group		
	3	5	10
Effect size (fold change)			
1.25	17 %	25 %	44 %
1.5	43 %	64 %	91 %
2	87 %	98 %	100 %
Sequencing depth (millions of reads)			
3	19 %	29 %	52 %
10	33 %	51 %	80 %
15	38 %	57 %	85 %

Example of calculations for the probability of detecting differential expression in a single test at a significance level of 5 %, for a two-group comparison using a Negative Binomial model, as computed by the RNASeqPower package

Artificial RNA spike-in

If it is important to you to accurately quantify absolute transcript concentration, you may want to consider artificial RNA, such as the ERCC spike-in standard which consists of 92 polyadenylated transcripts of varying lengths 250-2000 nucleotides and GC-content 51%. These RNA of known quantities can be used for the calibration of the RNA concentrations in each sample and to assess the sensitivity, coverage and linearity of your experiment, i.e., the overall technical performance.

Avoiding bias

The main goal of a well planned experiment is to improve the precision of the answers you will eventually get.

1. Identify the question of interest (What is the effect you are truly after?)
2. Attempt to identify possible sources of variability (nuisance factors)
3. Plan the experiment in a way that reduces the effect of the expected nuisance factors;
4. Protect yourself against unknown sources of variation.

Randomization

In addition to suffcient numbers of replicates, true randomization when selecting replicates and performing sample preparations can help to avoid unconscious selection bias that might be caused by sublte difference in the activity of the activity of the cells, their appearance, the growth pattern of cell lines etc. True randomization means: make the decision about any of the factors of interest by *tossing a coin*. This is fairly straight-forward when the factors are easily controlable, such as deciding which batches of cells to treat with a drug and which ones to keep as a control.

Blocking

Randomization is meant to protect you falling prey to spurious signals due to unintended batch effects. Usually, you will know about some factors that are very likely to be responsible for gene expression variation, such as sex, wheight, or the cell cycle status of your cells of interest. If it is feasible to group your samples into distinct classes (or “blocks") for these known sources of variation, a blocking experimental design may make sense and will help increase statistical sensitivity. For a blocking design, you will create complete sub-experiments for each class, i.e. all conditions of interest must be present in every block. By creating these blocks in which the nuisance factor is kept constant, you will be able to detect the changes that are due to the factor of interest without having to worry about the nuisance factor. If the blocking factor accounts for a sufficient amount of sample-to-sample variation, this will increase the sensitivity of the statistical tests to detect changes of interest – there is no guarantee though! Also, keep in mind though that within each block the assignment of treatments etc. should still be randomized.

Block what you can, randomize what you can not.

Introduction to our protocol

RNASeq technologies are required to become more sensitive, the goal being to detect rare transcripts in cells. However, sensitivity, accuracy and precision of transcript quantification strongly depend on how the mRNA is converted into the cDNA that is eventually sequenced. Especially when starting from low amounts of RNA, amplification is necessary to generate enough cDNA for sequencing. While it is known that PCR does not amplify all sequences equally well. It is unclear how PCR bias affects quantitative RNASeq analyses and to what extent PCR amplification adds noise and hence reduces the precision of transcript quantification. For detecting differentially expressed genes this is even more important than accuracy because it influences the power and potentially the false discovery rate.

The following RNASeq protocol is based on the protocol described in the paper *Characterization of directed differentiation by high-throughput single-cell RNASeq* by Soumillion et al. 2014 and taught to us by Dr. Rupert Oellinger of the Institute of Molecular Oncology and Functional Genomics TUM Munich. They established a modified protocol and suggested to improve it by elongating the BC from 6 to 8 and the UMI sequence from 10 to 12 bases, by doing that we should be able to use the original protocol again with some improvement in more distinct barcode and UMI regocnition.

The 3' digital gene expression (3'DGE) RNASeq protocol enables efficiect characterization of samples. It relies on a template-switching reverse transcriptase to convert poly(A)+ mRNA from isolated cells to cDNA decorated with universal adapters, well-specific barcodes and unique molecular identifiers (UMI's). This early barcoding allows all samples to be pooled right after reverse transcription . The primer sequences required for the library amplification are introduced at the 3'end during reverse transcription. Thus, PCR-duplicates in UMISeq data can always be identified via UMI. Decorated cDNA from multiple cells are then pooled, amplified and prepared for multiplexed sequencing using a modified transposon-based fragmentation approach that enriches for 3'ends and preserves strand information.

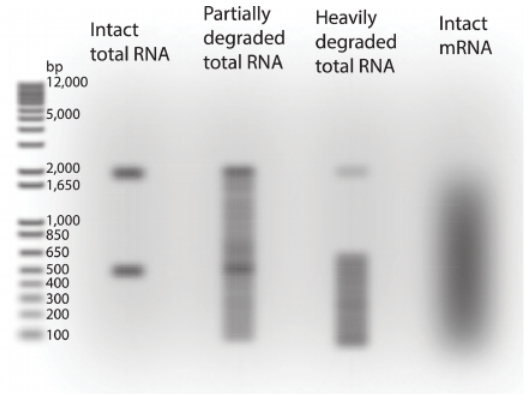
RNA extraction

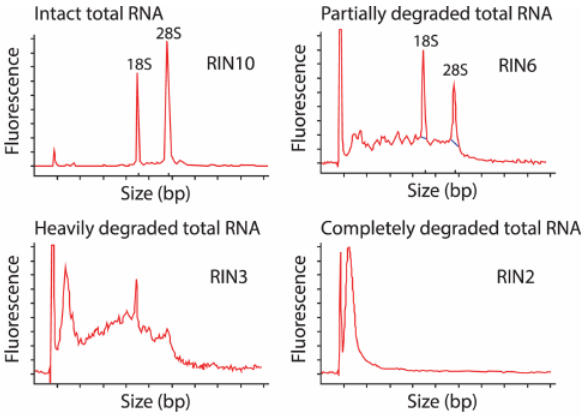
Before RNA can be sequenced, it must be extracted and separated from its cellular environment. The RNeasy Micro Kit (Cat No./ID: 74004) is used for that purpose. Automated RNA Purification from cells could be achieved with the Maxwell® RSC simplyRNA Cells Kit (AS1390).

Quality control RNA

RNA is much more susceptible to degradation than DNA and the quality of the extracted RNA molecules can strongly impact the results of the RNAseq experiment. Intact eukaryotic total RNA should yield clear

28S and 18S rRNA bands. Thw 28S rRNA band is approximately twice as intense as the 18S rRNA band. As RNA degrades, the 2:1 ratio of high quality RNA decreases, and low molecular weight RNA begins to accumulate. Agilent developed a software algorithm that allows for the calculation of an RNA Integrity Number (RIN) from a digital representation of the size distribution of RNA molecules (which can be obtained from an Agilent Bioanalyzer). The RIN number is based on a numbering system from 1 to 10, with 1 being the most degraded and 10 being the most intact.





Materials

Item	Source	Vendor
Maxima H Minus Reverse Transcriptase	EP0752	Thermo Fischer™
dNTPS	any	any
DNA Clean & Concentrator™	D4004	Zymoresearch
EDTA 0.5M	any	any
1N NaOH	any	any
Exonuclease I	M0293	NEB
KAPA HiFi HotStart PCR Kit	KK2501	Roche
Agencourt AMPure XP	A63880	Beckman Coulter
Qubit™ dsDNA HS Assay Kit	Q32854	Thermo Fisher™
Nextera® XT DNA Library Preparation Kit	FC-131-1024	Illumina®
NextSeq® 500/550 High Output v2 Kit (75 cycles)	FC-404-2005	Illumina®

Samples

In [0]:
```
#@title Libraries
# Libraries
import pandas as pd
from google.colab import drive, files
import io
```

In [0]:
```
#@title Format tables
# # Format tables printed to the screen
%load_ext google.colab.data_table
```

In [0]:

```
# mount google drive for data import
drive.mount('/gdrive', force_remount=True) # connect to google drive
%cd '/gdrive/My Drive/notebooks/RNASeq/DATA'
```

```
Go to this URL in a browser: https://accounts.google.com/o/oauth2/auth?client
_id=947318989803-6bn6qk8qdgf4n4g3pfee6491hc0brc4i.apps.googleusercontent.com&
redirect_uri=urn%3aietf%3awg%3aoauth%3a2.0%3aoob&response_type=code&scope=ema
il%20https%3a%2f%2fwww.googleapis.com%2fauth%2fdocs.test%20https%3a%2f%2fwww.
googleapis.com%2fauth%2fdrive%20https%3a%2f%2fwww.googleapis.com%2fauth%2fdri
ve.photos.readonly%20https%3a%2f%2fwww.googleapis.com%2fauth%2fpeopleapi.read
only (https://accounts.google.com/o/oauth2/auth?client_id=947318989803-6bn6qk
8qdgf4n4g3pfee6491hc0brc4i.apps.googleusercontent.com&redirect_uri=urn%3aiet
f%3awg%3aoauth%3a2.0%3aoob&response_type=code&scope=email%20https%3a%2f%2fww
w.googleapis.com%2fauth%2fdocs.test%20https%3a%2f%2fwww.googleapis.com%2faut
h%2fdrive%20https%3a%2f%2fwww.googleapis.com%2fauth%2fdrive.photos.readonly%2
0https%3a%2f%2fwww.googleapis.com%2fauth%2fpeopleapi.readonly)

Enter your authorization code:
··········
Mounted at /gdrive
/gdrive/My Drive/notebooks/RNASeq/DATA
```

In [0]:

```
#@title Sample Input
def sampleinput():
    """Ask's for Samples and add's them to a list."""
    prompt = "\nAdd a sample."
    prompt += "\nEnter 'quit' when you are done.\n"

    new_samples = []
    while True:
        sample = input(prompt)
        if sample == 'quit':
            break
        else:
            new_samples.append(sample)
    return pd.Series(new_samples)


def fileinput():
    """Reads a file in and creates a new dataframe."""
    uploaded = files.upload()
    try:
        df_sample_file = pd.read_csv(io.BytesIO(uploaded['samples.csv']))
        new_samples = df_sample_file['sample'].astype(str).values.tolist()
        return pd.Series(new_samples)
    except IOError:
        print("There is no file with that name.")


# ask for Sample input as manual, file or quit
while True:
    decision = input("\nHow do you want to enter your samples?\nManually -
    if decision == "1":
        new_samples = sampleinput()
        break
    elif decision == "2":
        new_samples = fileinput()
        break
    elif decision == "3":
        break
    else:
        # control for right input
        print("\nPlease choose option 1, 2 or 3.")
        continue

RNA_samples = new_samples
```

```
How do you want to enter your samples?
Manually - 1
File (named samples.csv) - 2
Quit - 3
2
```

Choose Files No file chosen

Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

```
Saving samples.csv to samples (1).csv
```

In [0]:

```
#@title 96-well format
# create a printable pipetting shema for 96-well plate format
import pandas as pd
shema = pd.read_csv('wellplate.csv') # read template in
RT_PCR = pd.DataFrame(shema[['column','row']])
RT_PCR['samples'] = RNA_samples # add the column with the samples
plate_shema = RT_PCR.pivot("column", "row", "samples")
plate_shema.index.name = None
plate_shema.columns.name = None
plate_shema
```

Out[6]:

	1	2	3	4	5	
A	DLD_1_siCtrl_1	DLD_1_siTROP2_1	DLD_1_siCtrl_2	DLD_1_siTROP2_2	DLD_1_siCtrl_3	DLD_1_siTI
B	HCT116_siCtrl_1	HCT116_siTROP2_1	HCT116_siCtrl_2	HCT116_siTROP2_2	HCT116_siCtrl_3	HCT116_siTI
C	SW480_R2_1	SW480_R2_2	SW480_R2_3	SW480_R2_4	SW480_R2_5	SW48
D	nan	nan	nan	nan	nan	
E	nan	nan	nan	nan	nan	
F	nan	nan	nan	nan	nan	
G	nan	nan	nan	nan	nan	
H	nan	nan	nan	nan	nan	

In [0]:

```
#@title Download pipetting shema
# transform the pandas df object in html and save pipetting_shema.html fil
pd.set_option('colheader_justify', 'center')   # FOR TABLE <th>

html_string = '''
<html>
  <head><title>HTML Pippeting Shema qPCR Plate</title></head>
  <link rel="stylesheet" type="text/css" href="df_style.css"/>
  <body>
    {table}
  </body>
</html>.
'''

# Save the as output as html.file
with open('pipetting_shema.html', 'w') as f:
    f.write(html_string.format(table=plate_shema.to_html(classes='mystyle'
files.download('pipetting_shema.html')
```

Reverse Transcription and Template Switch Reaction

Thermo Scientific Maxima H Minus Reverse Transcriptase (RT) was developed through in vitro evolution of M-MuLV RT. The enzyme possesses an RNA-dependent and DNA-dependent polymerase activity but lacks RNase H activity due to mutation in RNase H domain of M-MuLV RT. The engineered enzyme features dramatically improved thermostability, 50X higher processivity, robustness and increased synthesis rate compared to wild type M-MuLV RT.

The eliminated RNase H activity enables the enzyme to produce very long RNA transcripts up to 20 kb. Due to its high thermostability, the enzyme maintains full activity during the entire reverse transcription reaction and generates high yields of cDNA. The reaction temperature can be increased up to 65°C for efficient transcription of RNA regions with a high secondary structure or to improve specificity using gene specific primers. The extremely high processivity of Maxima H Minus enzyme results in increased resistance to common reaction inhibitors, such as guanidine, formamide and ethanol.

Note: Mix and briefly centrifuge all reagents after thawing, keep on ice.

Barcode plate

In [0]:

```
#@title Barcode plate
# barcode plate
barcodes_input = pd.read_csv('barcodes.csv') # read template in
barcodes = pd.DataFrame(barcodes_input[['column','row','barcode']])
barcode_plate = barcodes.pivot("column", "row", "barcode")
barcode_plate.index.name = None
barcode_plate.columns.name = None
barcode_plate
```

Out[9]:

	1	2	3	4	5	6	7	8
A	E3V6-A1-TAGCTTGT	E3V6-A2-GGCTACAG	E3V6-A3-TGTACCTT	E3V6-A4-TGCGATCT	E3V6-A5-TTGGTATG	E3V6-A6-TCATTGAG	E3V6-A7-TTACTCGC	E3V6-A8-TAGAACAC
B	E3V6-B1-CGATGTTT	E3V6-B2-CTTGTACT	E3V6-B3-TTCTGTGT	E3V6-B4-TAGTGACT	E3V6-B5-TGAACTGG	E3V6-B6-TGGCTCAG	E3V6-B7-TCGTTAGC	E3V6-B8-TCATCCTA
C	E3V6-C1-GCCAATGT	E3V6-C2-ACTTGATG	E3V6-C3-TCTGCTGT	E3V6-C4-TACAGGAT	E3V6-C5-TACTTCGG	E3V6-C6-TATGCCAG	E3V6-C7-TACCGAGC	E3V6-C8-TGCTGATA
D	E3V6-D1-ACAGTGGT	E3V6-D2-TGACCACT	E3V6-D3-TTGGAGGT	E3V6-D4-TCCTCAAT	E3V6-D5-TCTCACGG	E3V6-D6-TCAGATTC	E3V6-D7-TGTTCTCC	E3V6-D8-TAGACGGA
E	E3V6-E1-ATCACGTT	E3V6-E2-TGGTTGTT	E3V6-E3-TCGAGCGT	E3V6-E4-TGTGGTTG	E3V6-E5-TCAGGAGG	E3V6-E6-TACTAGTC	E3V6-E7-TTCGCACC	E3V6-E8-TGTGAAGA
F	E3V6-F1-GATCAGCG	E3V6-F2-TCTCGGTT	E3V6-F3-TGATACGT	E3V6-F4-TAGTCTTG	E3V6-F5-TAAGTTCG	E3V6-F6-TTCAGCTC	E3V6-F7-TTGCGTAC	E3V6-F8-TCTCTTCA
G	E3V6-G1-CAGATCTG	E3V6-G2-TAAGCGTT	E3V6-G3-TGCATAGT	E3V6-G4-TCGAAGTG	E3V6-G5-TCCAGTCG	E3V6-G6-TGTCTATC	E3V6-G7-TCTACGAC	E3V6-G8-TTGTTCCA
H	E3V6-H1-TTAGGCAT	E3V6-H2-TCCGTCTT	E3V6-H3-TTGACTCT	E3V6-H4-TAACGCTG	E3V6-H5-TGTATGCG	E3V6-H6-TATGTGGC	E3V6-H7-TGACAGAC	E3V6-H8-TGAAGCCA

1. Add reaction components into a sterile, nuclease-free 96well PCR plate on ice in the indicated order:

Reagent	Amount
Template RNA	5µl total RNA(5ng/µl)
barcoded Oligo(dT)	1µl 10µM E3V6NEXT(A3667)

E3V6NEXT: 5'-/5Biosg/ACACTCTTTCCCTACACGACGCTCTTCCGATCT[BC6]$N_{10}T_{30}$VN-3'
E5V6NEXT: 5'-iCiGiCACACTCTTTCCCTACACGACGCrGrGrG-3'

Note: Our primers have BC8 and N12 and do not have biotin. iC: iso-dC, iG: iso-dG, rG: RNA G.

1st strand cDNA

)

2. Prepare a master mix and add 4µl to every well of the plate for a total Volume of 10µl.

Reagent	Amount
5x RT-Buffer	2µl
10mM dNTPs	1µl
template switch oligo	0.1µl 100µM E5V6NEXT(A3667)

Reagent	Amount
Maxima RT	0.125µl
water	0.775µl
Total Volumne per well	10µl

In [0]:
```
#@title Download barcodes.tsv file
# Create a .tsv file for scripts later
df_barcodes = barcodes_input.copy()
df_barcodes['samples'] = RNA_samples
df_barcodes = df_barcodes[df_barcodes.samples != 'nan']
df_barcodes['8BC'] = df_barcodes['barcode'].str[-8:]
df_barcodes = df_barcodes.drop(df_barcodes.columns[[0, 1, 2]], axis=1)
df_barcodes
df_barcodes.to_csv('barcodes.tsv', index=False, header=False, sep='\t')
files.download('barcodes.tsv')
```

In [0]:
```
#@title Master mix calculation
# Get the sample number
count_nan = RNA_samples.value_counts()
if 'nan' in count_nan:
    samples = 96 - count_nan['nan']
else:
    samples = len(count_nan)

# Calculation of master mix depending an sample number + 10% to accomadate
print("To prepare a Master Mix for {} samples pipette the following:\n".fo
print("5x RT-Buffer  : ", round((int(samples) * 2 * 1.1),1))
print("10mM dNTPs    : ", round((int(samples) * 1 * 1.1),1))
print("100µM E5V6NEXT: ", round((int(samples) * 0.1 * 1.1),1))
print("Maxima RT     : ", round((int(samples) * 0.125 * 1.1),1))
print("Water         : ", round((int(samples) * 0.775 * 1.1),1))
```

```
To prepare a Master Mix for 32 samples pipette the following:

5x RT-Buffer  :  70.4
10mM dNTPs    :  35.2
100µM E5V6NEXT:  3.5
Maxima RT     :  4.4
Water         :  27.3
```

2nd strand cDNA

```
5'–ACACTCTTTCCCTACACGACGCGGG:cDNA:NB(A)30–3'
                            CCC:cDNA:NV(T)30(N)10[BC6]TCTAGCCTTCTCGCAGCACATCCCTTTCTCACA–5'
```

3. Incubate the plate in a pre-heated PCR machine at 42°C; for 90 min.

Resulting full length cDNA

```
5'–ACACTCTTTCCCTACACGACGCGGG:cDNA:NB(A)30(N)10[BC6]AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT–3'
3'–TGTGAGAAAGGGATGTGCTGCGCCC:cDNA:NV(T)30(N)10[BC6]TCTAGCCTTCTCGCAGCACATCCCTTTCTCACA–5'
```

cDNA clean-up

1. **Pooling**

First pool all RT reactions from the 96well plate into a 2ml Eppendorf tube. The clean-up is done using the DNA Clean & Concentrator™ Kit from Zymoresearch.

In [0]:
```
#@title After pooling:
print("You should have " + str(int(samples) * 10) + "µl in your tube.")
```

```
You should have 320µl in your tube.
```

The DNA Clean & Concentrator™-5 (DCC™-5) provides a hassle-free method for the rapid purification and concentration of high-quality DNA from PCR, endonuclease digestions, cell lysates, and other impure DNA preparations. It can also be used for post-RT cDNA clean-up and purification of sequencing-ready DNA from M13 phage. Simply add the specially formulated DNA Binding Buffer to your sample and transfer the mixture to the supplied Zymo-Spin™ Column. There is no need for organic denaturants or chloroform. Instead, the product features Fast-Spin column technology to yield DNA that is free of salts and contaminants in just 2 minutes. The purified DNA is ideal for DNA ligation, sequencing, labeling, PCR, microarray, transfection, transformation, and restriction digestion procedures.

2. **Hydrolysis**

- Add 2µl 0.5M EDTA and 2µl 1N NaOH to 10µl of RT reaction.

In [0]:
```
#@title Sample to reagents calculation:
print("To your " + str(int(samples) * 10) + "µl sample pool add "\
      + str(int(samples) * 2) + "µl 0.5M EDTA and " + str(int(samples) * 2
```

```
To your 320µl sample pool add 64µl 0.5M EDTA and 64µl 1N NaOH.
```

Note: The volumes of EDTA and NaOH should be scaled proportionally depending on the starting volume of the RT reaction.

- Incubate at 65°C for 15 minutes.

3. **Clean-up**

- Transfer the Hydrolysis reaction to a 15ml Falcon tube. Add 98µl (7 volumes) of **DNA Binding Buffer** to the hydrolysis reaction above and mix well .

In [0]:
```
#@title Binding buffer calculation:
print("Add " + str(7*(int(samples) * 10 + 4 * int(samples))) + "µl of DNA
```

```
Add 3136µl of DNA Binding Buffer.
```

Note: Neutralisation (pH) following RNA hydrolysis is not necessary as the DNA Binding Buffer will effectively neutralize the NaOH added to the reaction.

- Transfer 700µl mixture in the to a provided **Zymo-Spin™ Column** in a **Collection Tube**.

- Centrifuge for 30 seconds 10000g. Discard the flow-through. Repeat with another round of 700µl until all of the Binding mixture went through the column.

- Add 200µl **DNA Wash Buffer** to the column. Centrifuge for 30 seconds. Repeat the wash step.

- Add 18µl **DNA Elution Buffer** or water directly to the column matrix and incubate at room

temperature for one minute. Transfer the column to a 1.5ml microcentrifuge tube and centrifuge for 30 seconds to elute the DNA.

4. **Removal of residual primers**

Exonuclease I (*E.coli*) from NEB (M0393) is used for degradation of post-PCR primers. The enyzme catalyzes the removal of nucleotides from single-stranded DNA in the 3' to 5' direction.

Reagent	Amount
cDNA eluate	18µl
10x Reaction Buffer	2µl
Exonuclease I	1µl

Incubate the reaction at 37° for 30min followed by 80° for 20min to deactivate the enyzme.

STOPPING POINT **Products may be stored at 4°C.**

Amplification of cDNA with single primer PCR

Add the following MasterMix to the 20µl cDNA from step before.

Reagent	Amount
KAPA Hifi Hot Start	25µl
10µM SINGV6 primer	1µl
water	4µl
total volume	50µl

SINGV6: 5’-/5Biosg/ACACTCTTTCCCTACACGACGC-3’

Run the following PCR program.

Stage	Temperature	Time
Hold	98°C	3 minute
12 Cycles	98°C	15 seconds
.	65°C	30 seconds
.	68°C	6 minutes
Hold	72°C	10 minutes
Hold	4°C	up to 1 hour

Full length cDNA amplification:

Single primer PCR

```
                                                                3-'CGCAGCACATCCCTTTCTCACA-5'
5'-ACACTCTTTCCCTACACGACGCGGG:cDNA:NB(A)30(N)10[BC6]AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'
3'-TGTGAGAAAGGGATGTGCTGCGCCC:cDNA:NV(T)30(N)10[BC6]TCTAGCCTTCTCGCAGCACATCCCTTTCTCACA-5'
5'-ACACTCTTTCCCTACACGACGC-3'
```

Purification of PCR reaction

Full lenght cDNAs are purified with the Agencourt AMPure XP magnetic beads and quantified on the Qubit 2.0 Flurometer.

IMPORTANT!

- If performing this procedure in a hood with high air flow, turn off the ventilation to avoid excessive ethanol evaporation during bead clean-up.
- Incubate the AMPure^{TM}XP Reagent at room temperature for at least 30 minutes, then vortex thoroughly to disperse the beads before use. Pipet the solution slowly.

---

1. Briefly centrifuge the PCR stripe to collect the contents.

2. Use a 100µl pipette to measure the reaction volume in each tube. If the volume is <50µl, add nuclease-free water to bring the volume in each well to 50µl.

3. Add 30µl (0.6x sample volume) of Agencourt^{TM} AMPure^{TM} XP Reagent to each sample.

4. Vortex for 15 seconds, then incubate at room temperature for 5 minutes.

5. Transfer the stripe to the magentic stand, incubate for 5 minutes or until the solution clears, then remove and discard the supernatant without disturbing the pellet.

6. Wash the magnetic beads with freshly prepared 80% ethanol.

    - Add 150µl 80% ethanol to each sample.
    - Incubate at room temperature for 30 seconds until solution clears, then remove and discard the supernatant without disturbing the pellet.

      Note: Do not disturb the pellet or move the stripe side-to side between positions on the magnet during the wash steps.

7. Repeat step 7.

8. Keeping the stripe in the magnetic stand, use a smaller pipette (10µl) to remove any remaining ethanol, then air-dry the beads at room temperature for 4 minutes.

9. Add 20µl of 10mM Tris HCL, pH 7.5-8.5 or RSB buffer (Nextera Kit) to the pellet.

10. Vortex thoroughly to disperse the beads, incubate at room temperature for 5 minutes.

Transfer the stripe to the magnetic stand and incubate for at least 2 minutes, then transfer 18µl of the supernatant to new stripes.

Determination of DNA Concentration

Measure the DNA concentration with the HS dsDNA Kit using the Qubit instrument. The concentration should be around 1-2ng/µl.

STOPPING POINT **Products may be stored at 4°C.**

Tagmentation

The full-length cDNA was then used as input to the Nextera XT library preparation kit according to the manufacturers protocol, with the exeption that the i5 primer was replaced by the P5NEXTPT5.

```
5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG*A*T*C*T*-3′
```

where * = phosphorothioate bonds.

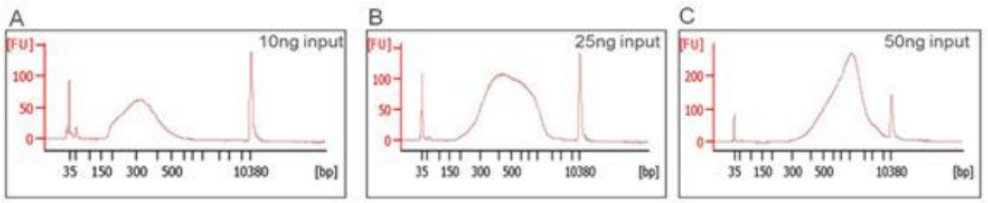
The phosphorothioate (PS) bond substitutes a sulfur atom for a non-bridging oxygen in the phosphate backbone of an oligo. This modification renders the internucleotide linkage resistant to nuclease degradation. Phosphorothioate bonds can be introduced between the last 3-5 nucleotides at the 5'- or 3'-end of the oligo to inhibit exonuclease degradation. Including phosphorothioate bonds throughout the entire oligo will help reduce attack by endonucleases as well.

Note The Nextera XT protocol is optimized for 1 ng of input DNA.Quantify the starting material before preparing libraries. Dilute starting material in RSB or 10 mM Tris HCl, pH 7.5–8.5.

The enzymatic DNA fragmentation used for this protocol is more sensitive to DNA input compared to mechanical fragmentation. Success depends on accurate quantification of input DNA. Use a fluorometric-based method to quantify input DNA. For example, if you use the Qubit dsDNA BR Assay system, use 2 µl of each DNA sample with 198 µl of the Qubit working solution. Avoid methods that measure total nucleic acid, such as NanoDrop or other UV absorbance methods.

The transposons are end point, in that they cleave double-stranded DNA one time only and then the step is complete. Therefore, successful tagmentation is highly dependent on mass.

- Using > 1ng can lead to undertagmentation of the sample
- Using < 1ng can lead to overtagmentation of the sample



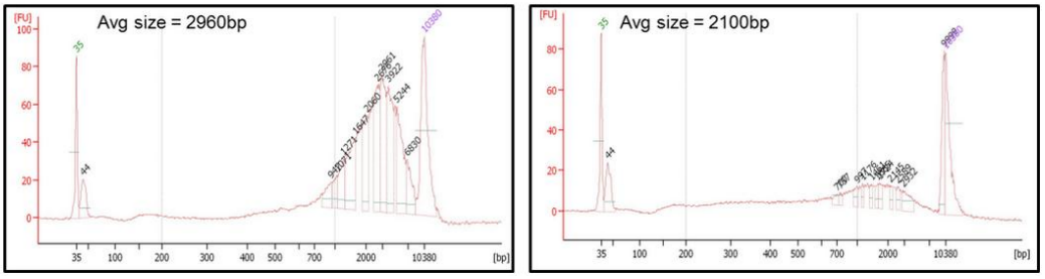
DNA Input	Median insert size		Mean Quality Metrics							
	On Bioanalyzer	Sequenced	Pass Filter	Align	Mismatch	>Q30	A+T Dropout	G+C Dropout	Diverse Reads	Duplicate Read Pairs
10ng*	180bp	N/A*	92%	64%*	0.50%	92%	not calculated			
25ng	330bp	230bp	93%	82%	0.50%	91%	4%	0%	2.53E+09	6.50%
50ng	1kb	346bp	92%	86%	0.60%	90%	6%	0%	2.67E+09	5%

A. 10 ng input library sequenced in triplicate
B. 25 ng input library sequenced in triplicate
C. 50 ng input library sequenced in triplicate

*A decrease in % align was observed in the run using 10 ng input due to significant sequencing into the adapter on this short-insert library. Therefore, the 10 ng input data set was not analyzed for drop out or diversity rates. Use shorter sequencing reads to mitigate this effect.

Undertagmentation

Final libraries above 1.2-1.5kb average size, including adapters, do not cluster on the flow cell.



These traces represent 2 Nextera libraries with undertagmented profiles. Most of the DNA in these libraries is greater than 1.5 kb in length, which does not efficiently form clusters on the flow cell.

Inhibitors of Enzymatic Reactions

Several factors can impair the performance of the Nextera enyzme:

- Proteins can coat DNA, preventing enzyme binding to the substrate
- Sequestration of enzyme cofactors by EDTA can negatively affect enzyme function.
- Proteinases, detergents, and phenol can degrade the enzyme.
- Change in ionic strength and pH caused by chemicals left over from libray preparation.

Best Practices

- Use careful sample handling with extraction protocols optimized to purify inhibitor-free nucleic acids.
- Apply accurate quantification approaches to detect inhibitors in the isolated DNA samples.
- Illumina recommends UV spectrophotometry for quality assessment and flourometric-based methods like Qubit for DNA quantification.
- When eluting or resuspending mucleic acids with water, make sure that the pH is 7.0-8.5.
- In the absence of a buffering agent, store the samples at 25°C to -15°C to prevent degradation.
- Make single-use aliquots of the input sample to prevent cross-contamination and avoid repeated freeze-thaw cycles.

In [0]:

```
#@title Best Practices for Nextera Library Prep Video
from IPython.display import HTML

# Youtube
HTML('<iframe width="1280" height="720" src="https://www.youtube.com/embed
```

Out[15]:

Assess DNA Quality

UV absorbance is a common method for assessing the quality of a DNA sample. The ratio of absorbance at 260 nm to absorbance at 280 nm is used as an indication of sample purity. This protocol is optimized for DNA with absorbance ratio values of 1.8–2.0, which indicates a pure DNA sample. Target a 260/230 ratio of 2.0– 2.2. Values outside this range indicate the presence of contaminants.

Overtagmentation

Samples that have been overtagmented by the transposons can lead to reduced libray yield, which can result in coverage dropout.

Causes of overtagmentation include:

- Inaccurate quantification that results in < 1ng for the tagmentation step
- Use of FFPE or degraded DNA samples
- Use of samller sized amplicons

Transposons require at least 300bp of genomic space to sit down on the DNA. For that reason, FFPE, degraded samples, and small amplicons are not supported.

How the Nextera XT Assay Works

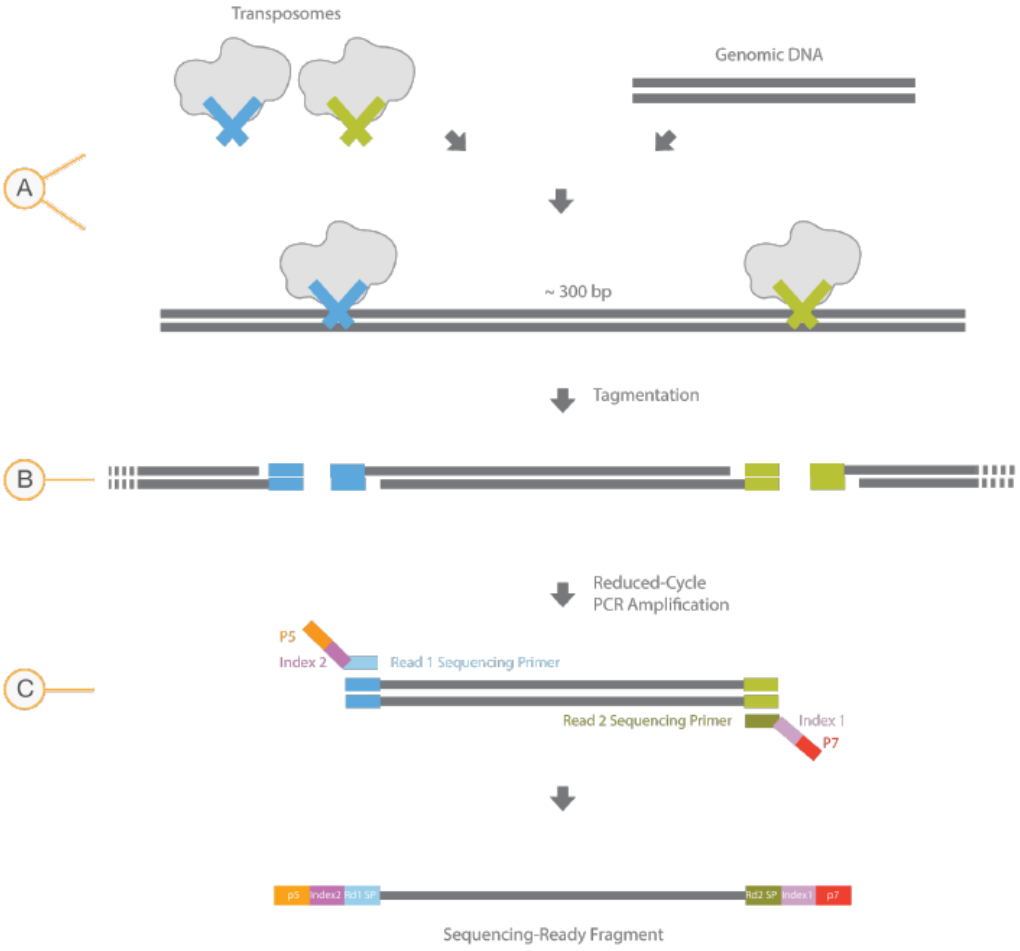
In [0]:
```
#@title Video
from IPython.display import HTML

# Youtube
HTML('<iframe width="1280" height="720" src="https://www.youtube.com/embed
```

Out[16]:

The Nextera XT DNA Library Prep Kit uses an engineered transposome to tagment cDNA, which is a process that fragments DNA and then tags the DNA with adapter sequences in one step. Limited-cycle

PCR uses the adapters to amplify the insert DNA. The PCR step also adds index adapter sequences on both ends of the DNA, which enables dual-indexed sequencing of pooled libraries on Illumina sequencing platforms.



Tagment cDNA

This step uses the Nextera transposome to tagment cDNA, which is a process that fragments DNA and then tags the DNA with adapater sequences in a single step.

Transposon based library (Nextera)

Tagmentation

```
5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT[BC6](N)10(T)30VN-Frag-3'
                                                3'-Frag-GACAGAGAATATGTGTAGAGGCTCGGGTGCTCTG-5'
```

Preparation

1. Prepare the following consumables:

Item	Storage	Instructions
cDNA	-25°C	Thaw on ice. Invert the thawed tubes 3-5 times, and then centrifuge briefly
ATM(Amplicon Tagment Mix)	-20°C	Thaw on ice. Invert the thawed tubes 3-5 times, and then centrifuge briefly
TD (Tagment DNA Buffer)	-20°C	Thaw on ice. Invert the thawed tubes 3-5 times, and then centrifuge briefly
NT (Neutralize Tagment Buffer)	RT	Check for precipitates. If present, vortex untill all particulates are resuspended.

Procedure

The Nextera XT protocol is optimized for 1ng of input DNA. Quantify the starting material before preparing libraries. Dilute starting material in molecular-grade water or 10mM Tris HCL, pH 7.5-8.5.

Use two reactions per library each with an input of 1ng cDNA.

1. Add the following volumes in the order listed to each pcr tube. Pipette to mix 10µl TD and 5µl normalized cDNA (0.20ng/µl).

2. Add 5µl ATM to each tube. Pipette to mix.

3. Centrifuge at 280xg at RT; for 15 seconds.

4. Place the stripe in the preprogramed thermal cycler and run the tagmentation program:

- choose the preheat lid option
- 55° for 5 minutes
- Hold at 10°

When the sample reaches 10°, **immediately** procced to the next step because the transposome is still avtive.

5. Add 5µl NT to each tube. Pipette to mix.

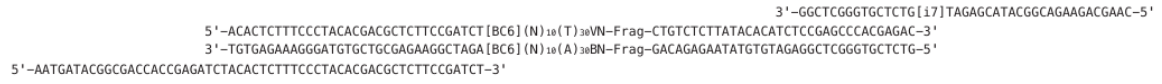
6. Centrifuge at 280xg at RT; for 15 seconds.

7. Incubate at room temperature for 5 minutes. The PCR plate contains 25µl tagmented and neutralized cDNA, all of which is used in the next step.

Amplify Libraries

This step amplifies the tagmented DNA using a limited-cycle PCR program. PCR adds the Index i7 and full i5/i7 adapter sequences to the tagmented DNA from the previous step. The adapters and Nextera PCR Master Mix are added directly to the 25µl of tagmented DNA from the previous step.

Library amplification (modified)



The adapters are required for cluster formation. Use the full amount of recommended input DNA and the specified number of PCR cycles, which helps ensure high-quality sequencing results.

Preparation

1. Prepare the following consumables:

Item	Storage	Instructions
P5NEXTPT5/i7_primer	-20°C	Only prepare adaptes being used. Thaw at room temperature for 20 minutes. Invert each tube to mix. Centrifuge briefly.
NPM (Nextera PCR Master Mix)	-20°C	Thaw on ice for 20 minutes.

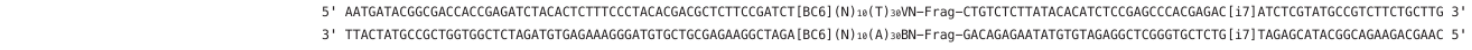
2. Save the following program on the thermal cycler:

Stage	Temperature	Time
Hold	72°C	3 minute
Hold	95°C	30 seconds
13 Cycles	95°C	10 seconds
.	60°C	30 seconds
.	72°C	60 seconds
Hold	72°C	5 minutes
Hold	4°C	up to 1 hour

Procedure

1. Add 5µl P5NEXTPT5 (A3669) to each tube.
2. Add 5µl i7_primer (A3733) to each tube.
3. Add 15µl NPM to each tube. Pipette to mix.
4. Centrifuge at 280xg at RT; for 15 seconds.
5. Place on the preprogrammed thermal cycler and run the PCR program. The volume is 50µl.
6. Pool the two reactions --> 100µl

Resulting library



STOPPING POINT **If you are stoping, store at 4°C for up to 2 days. Alternatively, leave on the thermal cycler overnight.**

Clean Up Libraries

This step uses AMPure XP beads to purify the library DNA and remove short library fragments.

Preparation

1. Bring Ampure XP Beads and RSB (Resuspensions Buffer) to room temperature.

2. Prepare fresh 80% ethanol from absolute ethanol.

Procedure

1. Centrifuge your samples at 280xg at 20°C for 1 minute.

2. Add 60µl AMPure XP beads to each well and pipette up and down 5 times.

3. Incubate at room temperature for 5 minutes.

4. Place on a magnetic stand and wait until the liquid is clear (~ 2 minutes).

5. Remove and discard all supernatant from each well.

6. Wash 2 times as follows.

- Add 200µl fresh 80% EtOH to each well.
- Incubate on the magnetic stand for 30 seconds.
- Remove and discard all supernatant from each well.

7. Using a 20µl pipette, remove residual 80% EtOH from each well.

8. Air-dry on the magentic stand for 5 minutes.

9. Remove from the magnetic stand and add 20µl RSB to each well and pipette 5 times up and down.

10. Incubate at room temperature for 2 minutes.

11. Place on a magnetic stand and wait until the liquid is clear (~ 2 minutes).

12. Transfer 18µl supernatant to a new tube.

STOPPING POINT **If you are stoping, store at -20°C for up to seven days.**

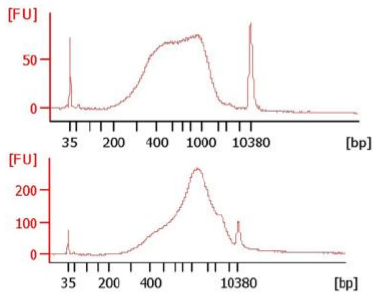
Nextera XT Quality Metrics

Two factors can cause cluster density fluctutations in libraries prepared with the Nextera XT DNA Library Kit:

- An average sample size that is too large or to small after tagmentation
- A final sample concentration that is too low

Check Library Size

Run 1µl of undiluted library on an Bioanalyzer. Typical libraries show a broad size distribution of ~250-1000 bp, look for a long low plateau. Various libraries can be sequenced with average fragment sizes as small as 250bp or as large as 1500bp.



These traces represent examples of a successful final library.

Results can vary due to the enzymatic shearing steps used in Nextera preps. As long as you have a distribution of sample from approximately 200 bp to 1.5 kb, you can proceed with sequencing.

Larger molecules cluster less efficiently than smaller molecules. If the fragment size after tagmentation is larger than expected, low cluster numbers are possible. The inverse is also true. The average expected library size after tagmentation is between 400bp and 1.2kb.

- **Short libraries indicate too little input DNA** Requantify the input DNA with a fluorometric method. Start with 10%-25% more input DNA. If the library peak is below 400bp and you want to continue with this library, dilute the library further.
- **Long libraries indicate too much input DNA or the presence of inhibitors** Start with less input DNA, make sure that the input DNA is free from inhibitors, and repear the quantification step.

Library Quantification

Measure library concentration using high sensitiviy dsDNA Qubit after library cleanup, and measure library size with a Bioanalyzer to calculate molarity.

- Bioanalyzer traces or qPCR are not acceptable methods for quantifiying Nextera libraries. Although a Bioanalyzer trace is a good method for assecing final library size, it is not accurate for quantification due to a wide sample size distribution.
- qPCR is appropriate only for libraries with a very narrow size range. A qPCR standard is not available for large size distribution of the final library.

Standard Normalization Calculation

1. After running the libraries on the Qubit and Bioanalyzer, the following formula is used to calculate the nM concentration of the library:

$$nM = (\frac{\text{Qubit ng/µl}}{\text{avg bp size x 660 g/mol}})x1000000$$

2. The library is then normalized to 2nM using:

$$C_i x V_i = C_r x V_r$$

then $V_r - V_i$ = µl resuspension buffer (RSB) to add

In [0]:

```
#@title Normalization calculation
Qubit =  9#@param {type:"number"}
avg_Basepairs_Agilent =  456#@param {type:"integer"}
nM = (Qubit / (avg_Basepairs_Agilent * 660)) * 1000000
vr = (nM * 2) / 2
rsb = round((vr - 2),2)
print("Take 2µl of your library and add {}µl RSB for a final concentration
```

```
Take 2µl of your library and add 27.9µl RSB for a final concentration of 2nM.
```

NextSeq Sequencing

In [0]:

```
#@title Video
from IPython.display import HTML

# Youtube
HTML('<iframe width="1280" height="720" src="https://www.youtube.com/embed
```

Out[18]:

To perform a sequencing run on the NextSeq 500, prepare a reagent cartridge and flow cell. Then follow the software prompts to set up and start the run. Cluster generation and sequencing are performed on-instrument. After the run, an instrument wash begins automatically using components already loaded on the instrument.

Cluster Generation

During cluster generation, single DNA molecules are bound to the surface of the flow cell, and then amplified to form clusters.

Sequencing

Clusters are imaged using two-channel sequencing chemistry and filter combinations specific to each of the fluorescently labeled chain terminators. After imaging of a tile on the flow cell is complete, the next tile is imaged. The process is repeated for each cycle of sequencing. Following image analysis , the software performs base calling, filtering, and quality scoring.

Analysis

As the run progresses, the control software automatically transfers base call (BCL) files to BaseSpace Sequence Hub, Local Run Manager, or another specified output location for secondary analysis.

Sequencing Run Duration

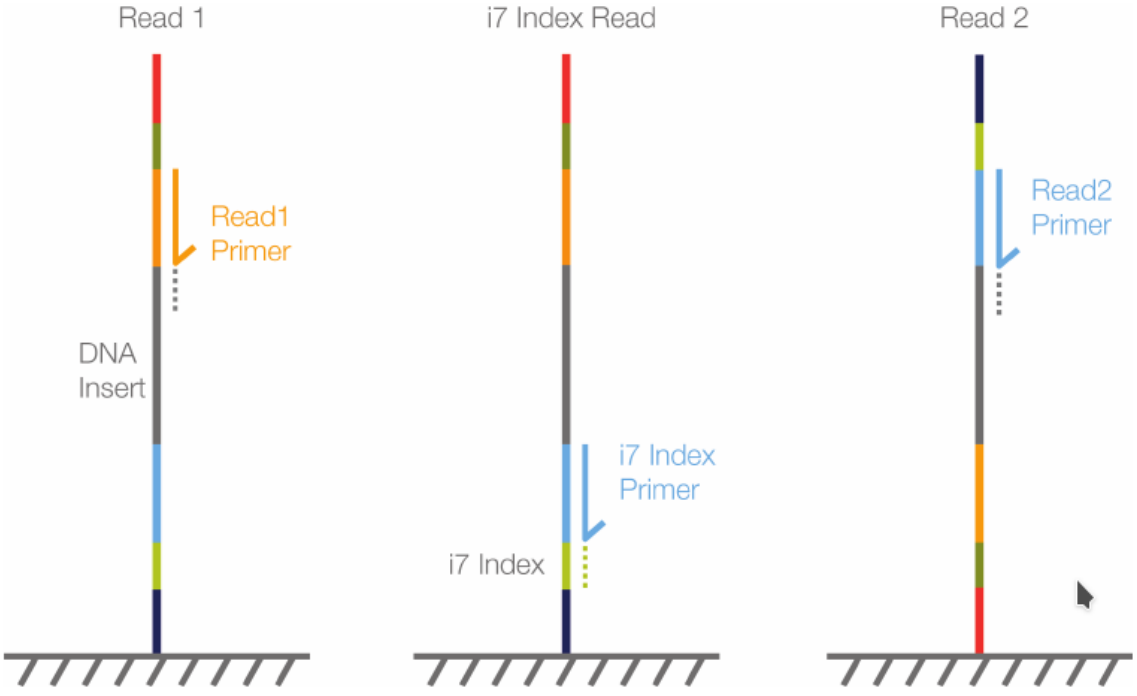
Sequencing run duration depends on the number of cycles performed. The maximum run length is a paired-end run of 150 cycles each read (2 x 150), plus up to eight cycles each for two index reads.

Number of Cycles in a Read

In a sequencing run, the number of cycles performed in a read is one more cycle then the number of cycles analyzed. For example, a paired-end 150-cycle run performs reads of 151-cycles (2 x 151) for a total of 302 cycles. At the end of the run, 2 x 150 cycles are analyzed. The extra cycle is required for phasing and prephasing calculations.

The NextSeq 500/550 High Output v2 Kit (75 cycles) is used for our RNA sequencing. The sequencing is similar to the Single-Indexed Sequencing.

Figure 1 Single-Indexed Sequencing



1. **Read 1** Read 1 with 20 cycles will be the Barcode (8) and UMI (12). The Read follows the standard Read 1 protocol using SBS reagents. The Read 1 sequencing primer is annealed to the template strand during the cluster generation step.
2. **Index Read preparation** The Read 1 product is removed and the Index 1 (i7) sequencing primer is annealed to the same template strand, producing the Index 1 (i7) Read.
3. **Index 1 (i7) Read** Following Index Read preparation, the 7-cycle Index 1 (i7) Read is performed.
4. **Read 2 resynthesis** The Index Read product is removed and the original template strand is used to regenerate the complementary strand. Then, the original template strand is removed to allow hybridization of the Read 2 sequencing primer.
5. **Read2** Read 2 which is the cDNA read follows the standard paired-end sequencing protocol using SBS reagents. It will be 75 cycles long.

Denature and Dilute Library

Protocol A (Standard Normalization Method) of the Illumina NextSeq System Denature and Dilute Libraries Guide is used.

Note Typically, it is important that not more then 1mM NaOH is in the final solution after diluting with HT1. However, introducing 200mM Tris-HCL ensures that the NaOH is fully hydrolyzed in the final solution. As a result, template hybridization is not affected even if the final NaOH concentration with HT1 is greater then 1mM.

Prepare Reagents

Prepare a fresh Dilution of NaOH

1. Prepare a fresh dilution of NaOH by combining the following volumes in a microcentrifuge tube:
2. Laboratory-grade water (800µl)
3. Stock 1.0N NaOH (200µl)
4. Invert the tube several times to mix and use the fresh dilution within **12 hours**.

The result is 1ml of 0.2N NaOH.

Prepare HT1

1. Remove HT1 from -20°C storage and thaw at room temperature.
2. Store at 2°C to 8°C until you are ready to dilute denatured libraries.

Prepare RSB

Note In place of RSB, you can use 10mM Tris-HCl, pH 8.5 with 0.1% Tween 20.

1. Remove the tube RSB from -20°C storage and thaw at room temperature.
2. Store thawed RSB at 2°C to 8°C until you are ready to dilute libraries.

Denature Libraries

1. Combine the following volumes of library and freshly diluted 0.2N NaOH in a microcentrifuge tube.

Starting Library Concentration	Library	0.2N NaOH
4nM	5µl	5µl
2nM	10µl	10µl
1nM	20µl	20µl
0.5nM	40µl	40µl

2. Vortex briefly and then centrifuge at 280xg for 15 secs.
3. Incubate at room temperature for 5 minutes.
4. Add the following volume of 200mM Tris-HCl, pH 7.

Starting Library Concentration	200mM Tris-HCl, pH 7
4nM	5µl
2nM	10µl
1nM	20µl
0.5nM	40µl

5. Vortex briefly and then centrifuge at 280xg for 15 secs.

Dilute Denatured Libraries to 20pM

1. Add the following volume of prechilled HT1 to the tube of denatured libraries.

Starting Library Concentration	Prechilled HT1
4nM	985µl
2nM	970µl
1nM	940µl
0.5nM	880µl

The result is 20pM denatured library.

2. Vortex briefly and then centrifuge at 280xg for 15 secs.
3. Place the 20pM libraries on ice until you are ready to proceed to final dilution.

Dilute Libraries to Loading Concentration

High Output Kits

Dilute the denatured 20pM library solution to 1.8pM as follows.

- Denatured library solution 117µl
- Prechilled HT1 1183µl
- Invert to mix and then pulse centrifuge.

The total colume is 1.3ml at 1.8pM.

The **Rad protocol** uses different numbers to generate a cluster density ~180 cluster/mm^2:

Dilute the denatured 20pM library solution to 1.4pM as follows.

- Denatured library solution 90µl
- Prechilled HT1 1210µl
- Invert to mix and then pulse centrifuge.

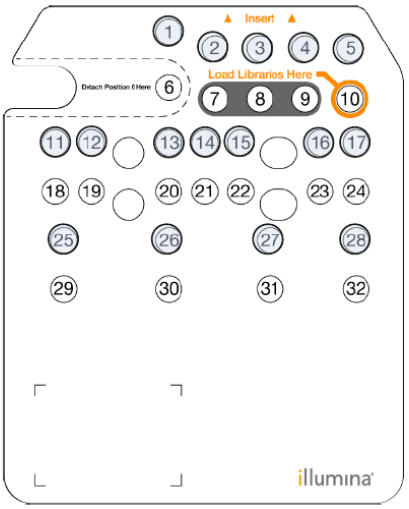
The total volume is 1.3ml at 1.4pM.

NextSeq 500/550 Reagent Cartridge Preparation

1. Thaw cartridge in RT water bath for 60 to 90 min
    - Do not submerge above lower lid edge
    - Ensure all reagents are fully thawed
2. Manually invert cartridge 5x
3. Tap cartridge on hard surface to remove bubbles and dislodge water from cart
4. Dry the base of the cartridge

NextSeq 500/550 Reagent Cartridge v2 Overview

Position(s)	Reagent Description
1-5	Wash Reservoir
6	Denaturation reagent
7-9	Reserved for custom primers
10	Library loading reservoir
11	0.12% Bleach
12-17	Wash Reservoir
18	Conditioning reagent
19	Resynthesis mix
20	Read 1 sequencing primer



Position(s)	Reagent Description
21	Read 2 sequencing primer
22	Index sequencing primers
23	Read 1 linearization mix
24	Read 2 linearization mix
25-28	Wash Reservoir
29	Cleavage mix
30	Scan mix
31	Incorporation mix
32	Amplification mix

Optional: NextSeq Custom Primers

This step is not needed in our case, because the sequencing primers are already included in the cartridge (20-21 position).

Prepare and Add Custom Primers

Make sure that the NextSeq reagent cartridge is thawed and inspected before proceeding.

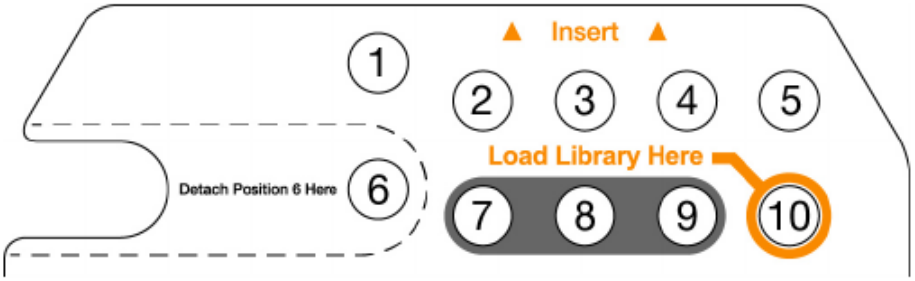
Prepare Custom Primers

1. If frozen, thaw each custom primer to be used.
2. Use HT1 to dilute each custom primer to yield 2 ml at 0.3µM final concentration (6µl 100µM in 2ml HT1)

Add Custom Primers to Reagent Cartridge

1. Use a low-lint lab tissue to wipe clean the foil seal covering each position associated with the custom primer.

Figure 1 Custom Primer Positions



Postion	Custom Primer	Name	Sequence	Cycles
7	Custom Read 1 primer	TruSeq-R1seq	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	20
8	Custom Read 2 primer	Nextera Read 2	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	75

2. With a clean pipette tip, pierce the foil seal covering each postion associated with the custom primer.
3. Add the 2ml of custom primer in the appropriate position on the reagent cartridge. Avoid touching the foil seal as you dispense the primer.

Setup Using NCS in Standalone Configuration

1. On the NCS Run Setup screen, select the checkbox for each custom primer to be used with the run. During the run , the software directs the sippers to pull primers from the appropriate reservoir.
2. When you have finished specifying run parameters, select **Next**.

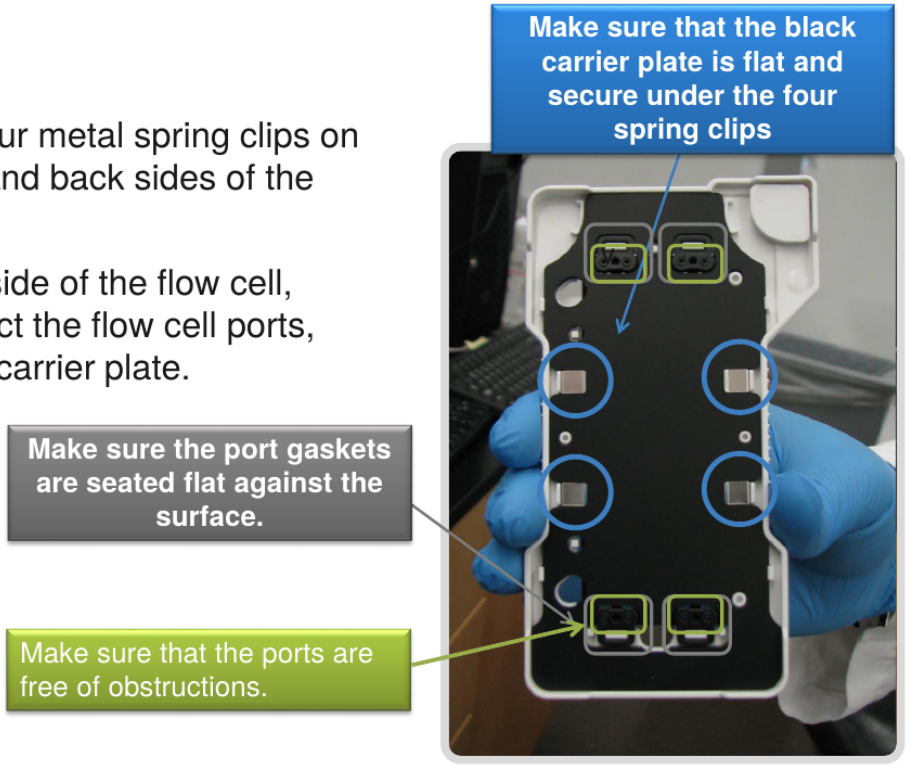
Prepare the Flow Cell

1. Leave the flow cell package at room temperature for 30 minutes before use
2. Remove the foil packaging and then remove the flow cell cartridge from the plastic clamshell casing
3. Clean the glass surface of the flow cell using an alcohol wipe if neccessary
4. Dry the glass with a lint-free tissue or lens paper

Note: After the foil packaging has been opened, use the flow cell within the next 12 hours.

-

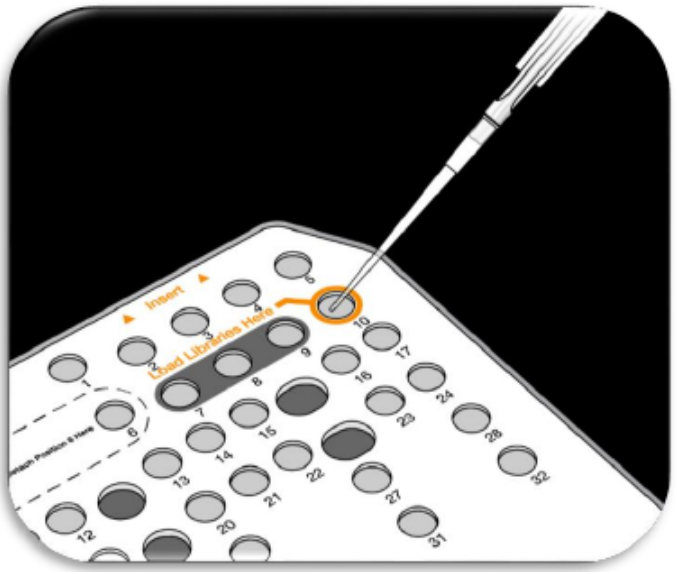
- Inspect the four metal spring clips on both the top and back sides of the flow cell
- On the back side of the flow cell, visually inspect the flow cell ports, gaskets, and carrier plate.



Load Libraries into Cartridge

1. Load sample in highlighted postition on reagent cartridge

    - Clean the foil seal covering reservoir 10 labeled **Load Library Here** using a low-lint tissue.
    - Pierce the seal with a clean 1ml pipette tip.
    - Add sample by pipetting to bottom of the well 1.3ml library
    - Avoid air bubbles



Start the Run in NCS

1. Form the Home screen, select **Experiment**.
2. On the Select Assay screen, select **Sequence**. The Sequence command opens the imaging compartment door, releases consumables from a previous run, and opens the series of run setup screens. A short delay is normal.

Run Modes

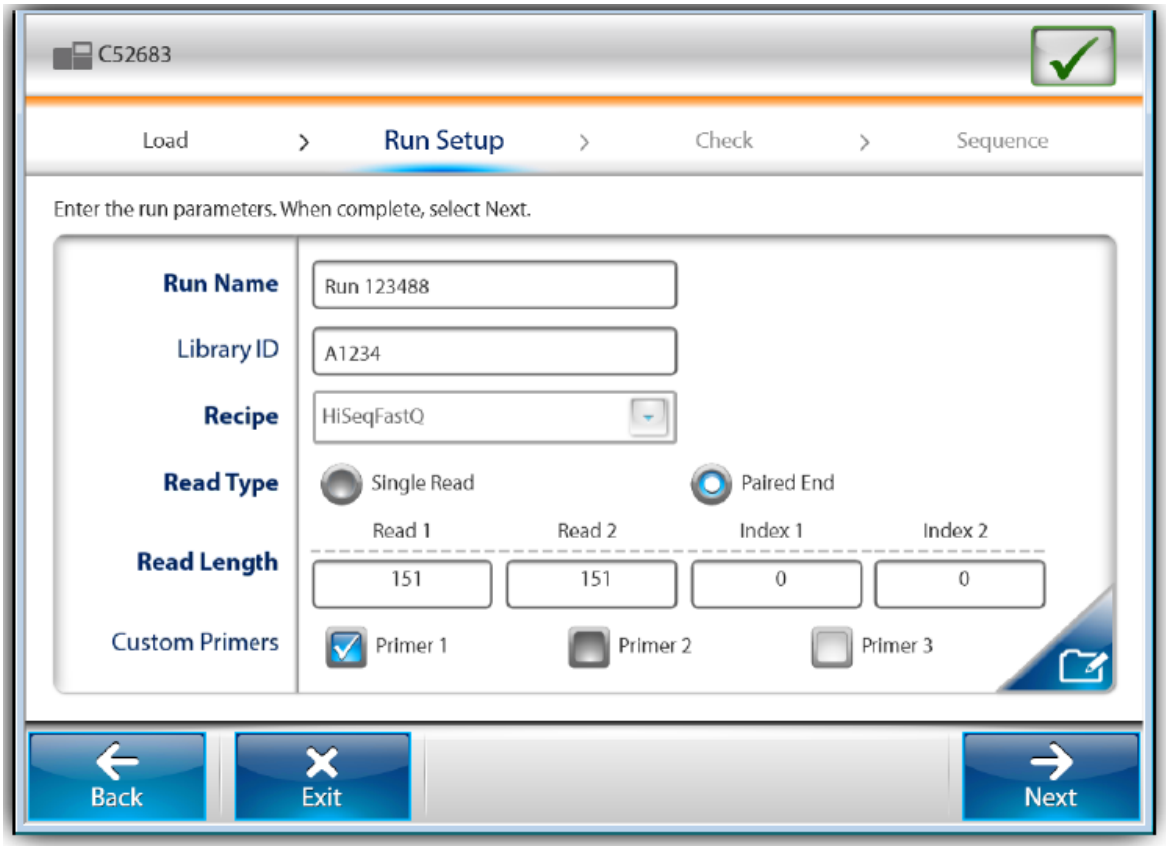
When setting up a sequencing run, you select one of the following run modes to determine where to enter run information and how to analyze data.

Run Mode	Run Information	Data Analysis
Local Run Manager	Enter in Local Run Manager.	The software saves data to specified output folder for automatic analysis in Local Run Manager.
Manual	Enter in the NCS.	The software saves to a specified output folder for later analysis off the instrument.

Enter Run and Analysis Parameters in NCS (Manual Run Mode)

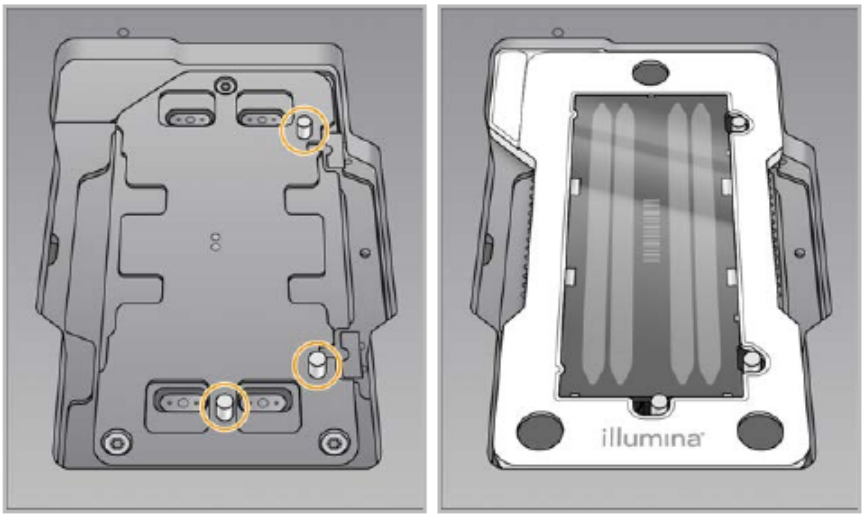
1. Enter a run name of your preference.
2. **[Optional]** Enter a library ID of your preference.
3. From the Recipe drop-down list, select a recipe. Only compatible recipes are listed.
4. Select a read type, either Single Read or Paired End.
5. Enter the number of cycles for each read in the sequencing run.
    - Read 1 - Enter 20 cycles
    - Read 2 - Enter 75 cycles
    - Index 1 - Enter the number of cycles required for the Index 1 (i7) primer.
    - Index 2 - we do not have a i5 Index The control software confirms your entries using the following criteria:
    - Total cycles do not exceed the maximum cycles allowed
    - Cycles for Read 1 are greater than the 5 cycles used for template generation
    - Index Read cycles do not exceed Read 1 and Read 2 cycles

6. **[Optional]** If you are using custom primer, select the checkbox for the primers used.
    - Read 1 - Custom primer for Read 1
    - Read 2 - Custom primer for Read 2
    - Index 1 - Custom primer for Index 1
    - Index 2 - Custom primer for Index 2
    - Output folder location - Change the output folder location for the current run. Select **Browse** to navigate to a network location.
    - Sample Sheet - Select **Browse** to navigate to a sample sheet (optional).
7. Select **Next**.
8. **[Optional]** Select the Edit icon to change run parameters.
9. Select **Next**.



Load the Flow Cell

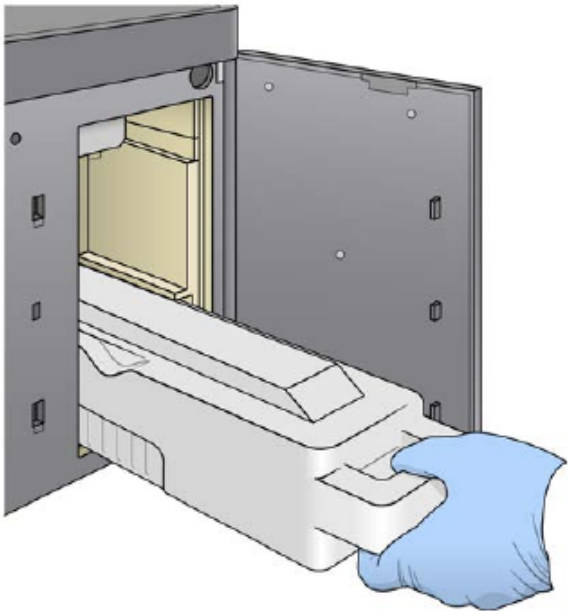
1. Remove the used flow cell from a previous run.
2. Align the flow cell over the alignment pins and place the flow cell on the stage.



3. Select **Load**. The door closes automatically, the flow cell ID appears on the screen, and the sensors are checked.
4. Select **Next**.

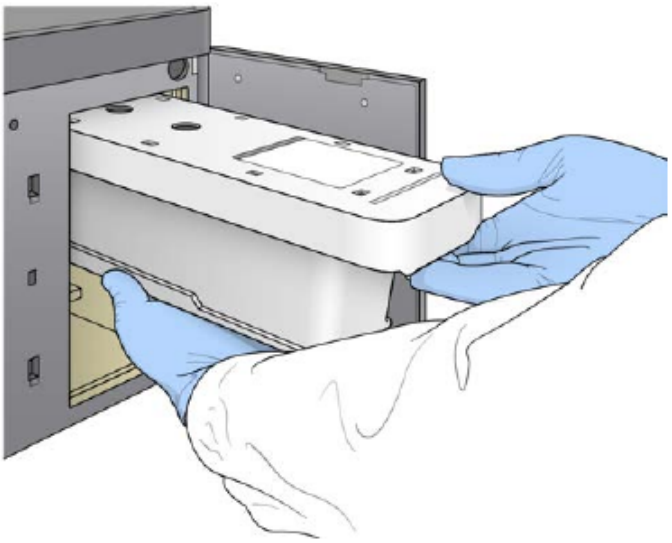
Empty the Spent Reagents Container

1. Remove the spent reagents container and discard the contents in accordance with applicable standards.
2. Slide the empty spent reagents container into the bufer compartment until it stops. An audible click indicates that the container is in postion.



Load the Buffer Cartridge

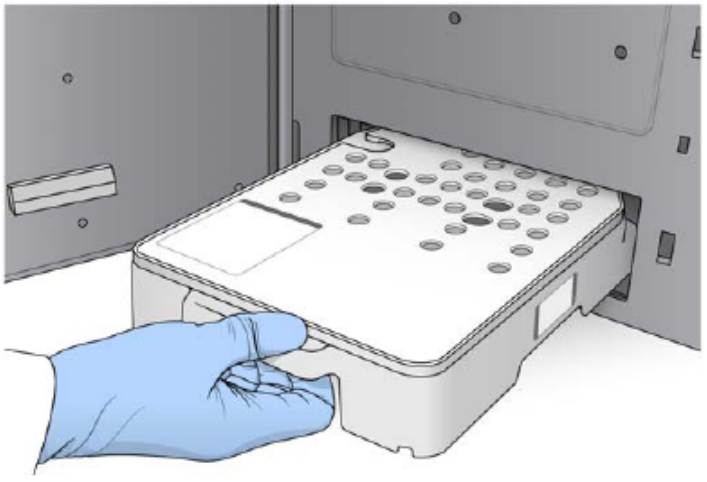
1. Remove the used buffer cartridge from the upper compartment.
2. Slide a new buffer cartridge into the buffer compartment until it stops. An audible click indicates that the cartridge is in position, the buffer cartridge ID appears on the screen, and the sensor is checked.



3. Close the buffer compartment door, and select **Next**.

Load the Reagent Cartridge

1. Remove the used reagent cartridge from the reagent compartment. Dispose of unused contents in accordance with applicable standards.
2. Slide the reagent cartridge into the reagent compartment until the cartridge stops, and then close the reagent compartment doorr.



3. Select **Load**. The software moves the cartridge into postion automatically (~30 seconds), the reagent cartridge ID appears on the screen, and the sensors are checked.
4. Select **Next**.

Review Automated Check

The software performs an automated check of the system. During the check, the following indicators appear on the screen:

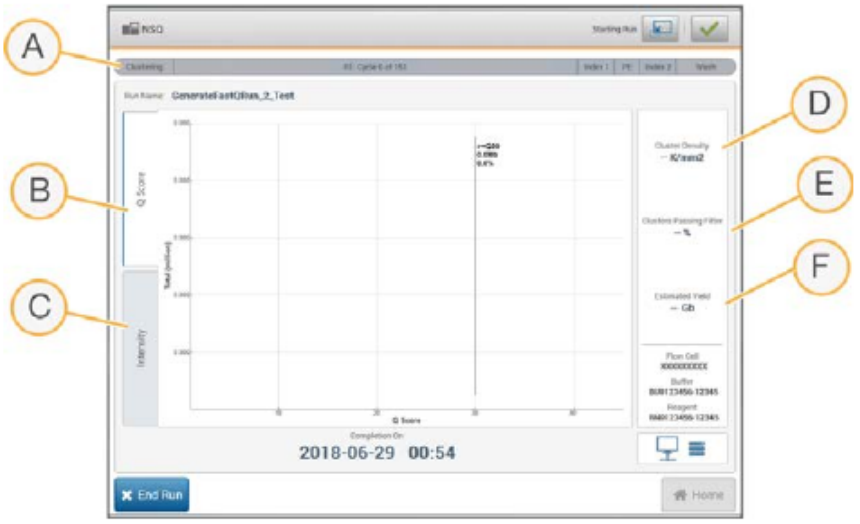
- Gray checkmark - The check has not been performed yet.
- Progress icon - The check is in progress.
- Green checkmark - The check passed.
- Red - The check did not pass. For any items that do not pass, an action is required before you can proceed.

Start the Run

When the automated check is complete, select **Start**. The sequencing run begins.

Monitor Run Progress

Monitor run progress, intensities, and quality scores as metrics appear on the screen.



- A **Run progress** - Shows the current step and number of cycles completed for each read. The progress bar is not proportional to the run rate of each step. Use the time remaining in the upperright corner to determine actual duration.
- B **Q-score** - Shows the distribution of quality scores (Q-scores).
- C **Intesity** - Shows the value of cluster intensities of the 90th percentile for each tile. Plot colors indicate each base: red is A, green is C, blue is G, and black is T. Colors match base indicators used in the Sequencing Analysis Software.
- D **Cluster Density (K/mm^2)** - Shows the number of clusters detected for the run.
- E **Clusters Passing Filter (%)** - Shows the percentage of clusters passing filter.
- F **Estimated Yield (Gb)** - Shows the number of bases projected for the run.

Cycles for Run Metrics

Run metrics appear at different points in a run.

- During the cluster generation steps, no metrics appear.
- The first five cycles are reserved for template generation.
- Run metrics appear after cycle 25, including cluster density, clusters passing filter, yield, and quality scores.

Data Transfer

Depending on the analysis configuration selected, an icon appears on the screen during the run to indicate the data transfer status.

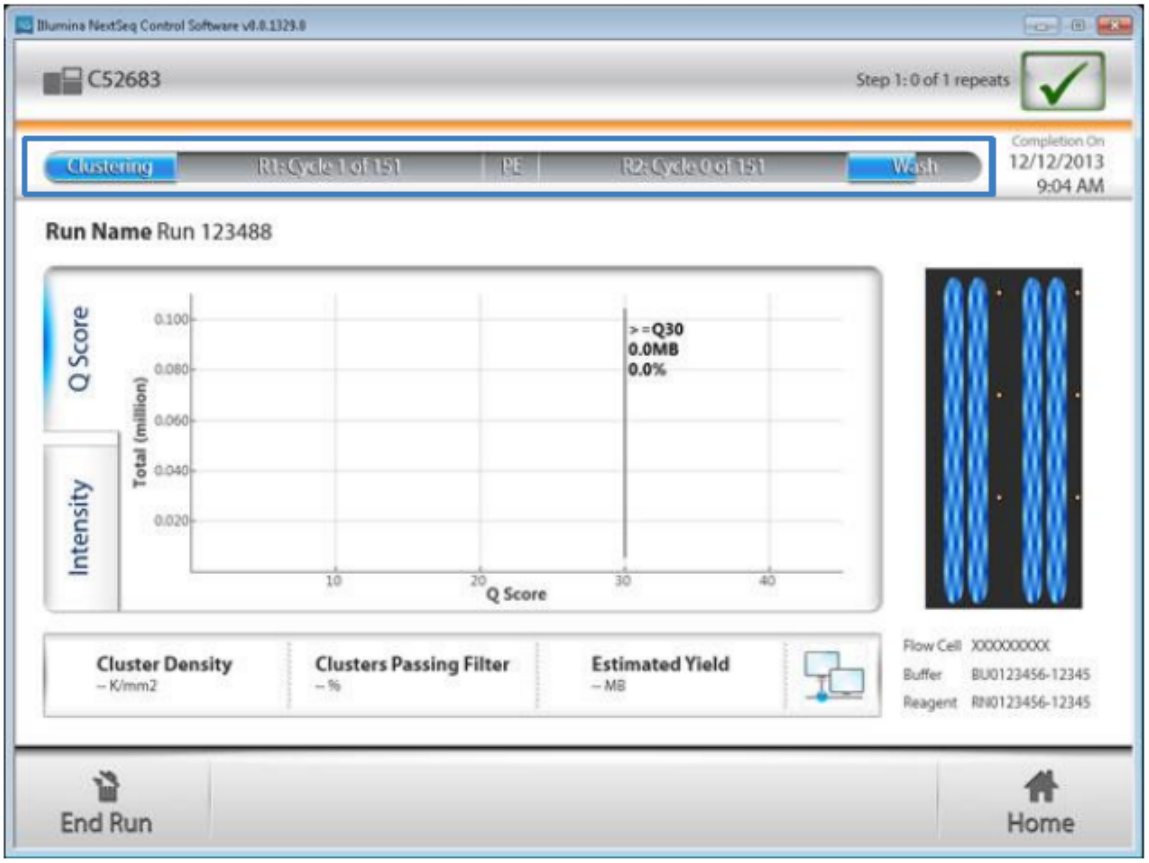
Status	Local Run Manager	Output Folder	Illumina BaseSpace
Connected			
Connected and transferring data			
Disconnected			
Disabled			

If data transfer is interrupted during the run, data are stored temporarily on the instrument computer. When the connection is restored, data transfer resumes automatically. If the connection is not restored before the run ends, manually remove the data from the instrument computer before a subsequent run can begin.

Post-Run Procedures

Post-Run Wash

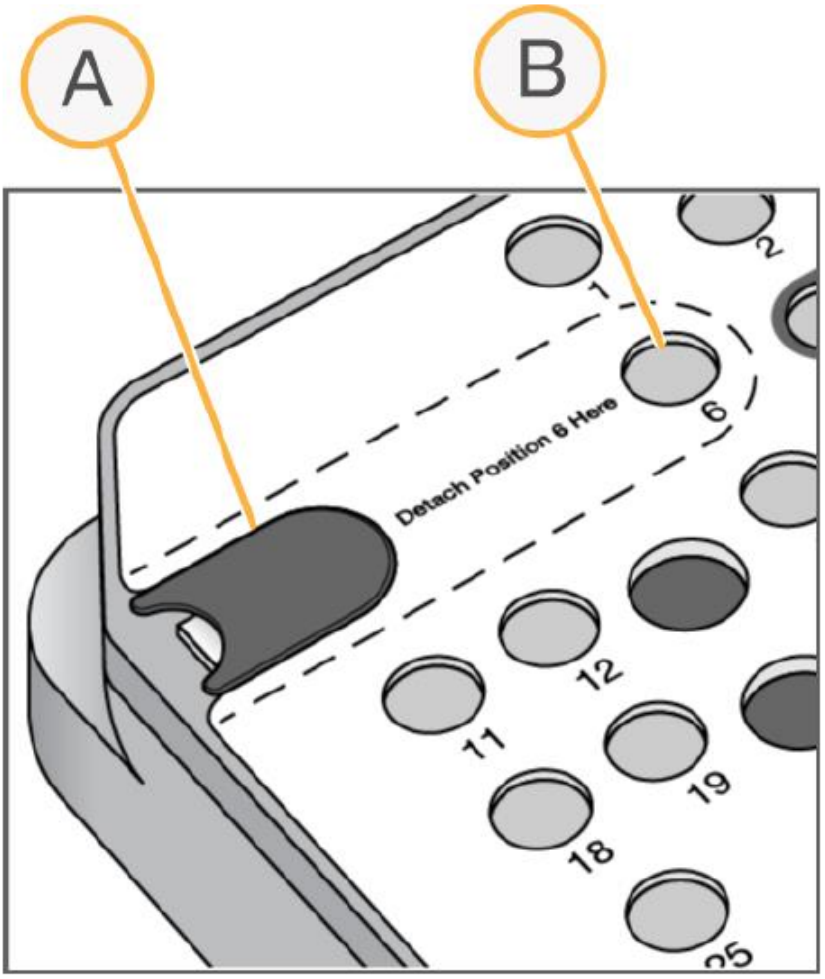
When the sequencing run is complete, the software initiates an automatic post-run wash. The post-run wash uses wash solution provided in the buffer cartridge and NaOCl provided in the reagent cartridge. The automatic post-run wash takes approximately 90 minutes. When the wash is complete, the Home button becomes active. Sequencing results remain visible on the screen during the wash. After the wash, the sippers remain in the down position to prevent air from entering the system. Leave the cartridge in place until the next run.



Expected Wash Deliveries	Approximate Run Time
17.25 ml	90 minutes

Removable Reservoir

- Position 6 of the reagent cartridge include a denaturation reagent (LDR) that contains formamide
- The removable reservoir in position 6 allows for safe disposal of any unused LDR after the sequencing run



A. Protective removable tab
B. Position 6

Manual Instrument Wash

- Perform manual instrument wash if instrument has been idle for 14 days or more
- Always use a wash solution of laboratory-grade water mixed with Tween 20 (0.05% Tween 20 final concentration)

Quick Wash

- Duration: 20 minutes
- Flushes system with wash buffer
- Use when instrument has been idle for 14 days or more or when restarting after instrument shut down

Manual Post-Run Wash

- Duration: 90 minutes
- Flushes system with wash buffer plus 0.12% NaOCl added to reagent wash cartridge
- Use if run was interrupted and automatic post-run wash was not performed

---

On-instrument Disk Space

- The NextSeq system only has enough disk storage to hold data from one complete run
- Best practise is to save sequencing run to BaseSpace or a network drive
- Ensure that the any run data from has been cleared from the instrument hard drive

In [0]:
```
1
```