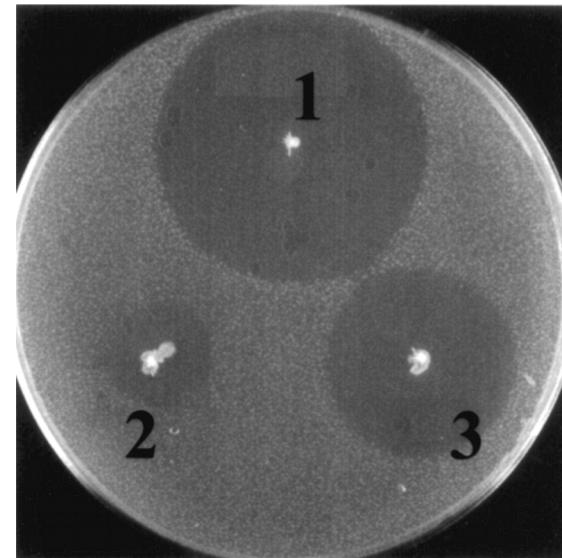


The biological question / hypothesis is the most important thing

What's the question/hypothesis underlying our experiment?

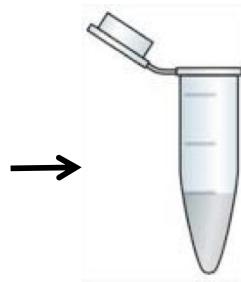


Is there a better/cheaper approach? Why sequencing?

You should think through every step of the process before you begin



Samples



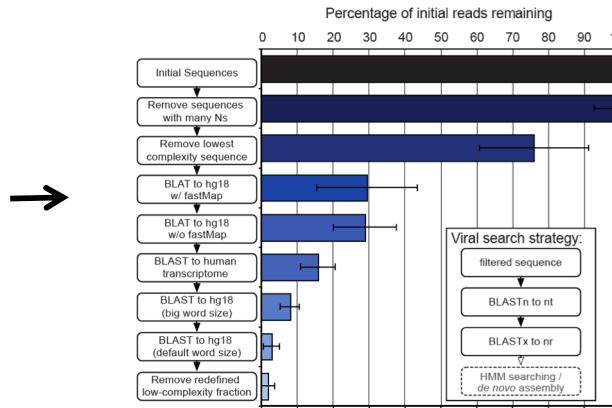
Nucleic acid



Library prep



Sequencing



Analysis



Follow-up

A number of questions related to sample processing and sequencing

- What type of samples will you start with?
- Will you pre-process your samples prior to nucleic acid extraction?
- What type of nucleic acid will you extract?
- Will you do enrichment (e.g. exome capture) or subtraction (e.g. ribosomal depletion) to focus on a subset of the nucleic acid? Is this necessary?
- What library prep protocol will you follow?
- How many samples will you be sequencing?
- How many replicates do you need? (more about replicates later)
- How many reads do you need for each sample?
 - For genome assembly or variant detection: how much coverage do you need?
 - For quantification expts, how well do you want to be able to sample low abundance molecules?
- What will be the impact of errors on your analysis?
- What type of sequencing?
- Paired end vs. single end
- Read length – short, longer, super long? A combination?
- What platform? Illumina, PacBio, Nanopore, other?

<https://genohub.com/next-generation-sequencing-guide/>

The screenshot shows a web browser window with the URL <https://genohub.com/next-generation-sequencing-guide/> in the address bar. The page has a red header with the genohub logo and navigation links for 'Find Genomics Services', 'Resources', and 'Sign In'. A sidebar on the left lists various sequencing parameters: Overview, Type of Run, Read Length, Number of Reads, Depth of Coverage (DNA), Depth of Coverage (RNA), Replication, Randomization and Multiplexing, Poor Quality Sequencing Run, Library Preparation, and Custom Sequencing Primers. A green button at the bottom of the sidebar says 'Search for NGS Services'. The main content area features a large heading 'Designing Your Next Generation Sequencing Run'. Below it, a section titled 'Type of Run – Single Read (SR) or Paired End (PE)' provides information about the two sequencing types, their benefits, and applications like RNA-Seq and ChIP-Seq.

Secure | <https://genohub.com/next-generation-sequencing-guide/>

Coloradoan FoCo W blastn blastx blastp GenBank Denver Post Gallery · mbostock/... Website_Jekyll

genohub Find Genomics Services Resources Sign In

Designing Your Next Generation Sequencing Run

[Back to NGS Handbook](#)

While we've standardized the way sequencing runs are ordered across all platforms, customization is required when determining the run parameters to achieve your experimental goals. If you're new to high throughput sequencing and have questions about how you should design your sequencing run, fill out our [free consultation form](#) and we'll get in touch with you to help.

We highly recommend that use [Genohub's NGS Matching Engine](#) as a great tool to determine the right amount of sequencing capacity on various instruments and easily explore different options. Simply [enter your specifications](#) and instantly see services that match your output requirements.

Type of Run – Single Read (SR) or Paired End (PE)

With single read runs the sequencing instrument reads from one end of a fragment to the other end. Paired end runs read from one end to the other end, and then start another round of reading from the opposite end. Single read runs are faster, cheaper and are typically sufficient for profiling or counting studies such as [RNA-Seq](#) or [ChIP-Seq](#).

Paired end runs give additional positioning information in the genome, making it a good choice for de novo genome assembly as well as making it easier to resolve structural re-arrangements such as deletions, insertions and inversions. Experiments designed to study splice variants, epigenetic modifications (methylation) and SNP identification are best served by paired-end runs. While paired end runs are more costly and time consuming, you get back twice the amount of data at less than double the cost to sequence.

Overview

Type of Run

Read Length

Number of Reads

Depth of Coverage (DNA)

Depth of Coverage (RNA)

Replication, Randomization and Multiplexing

Poor Quality Sequencing Run

Library Preparation

Custom Sequencing Primers

Search for NGS Services

<https://genohub.com/next-generation-sequencing-guide/>

Recommended RNA-Seq Parameters

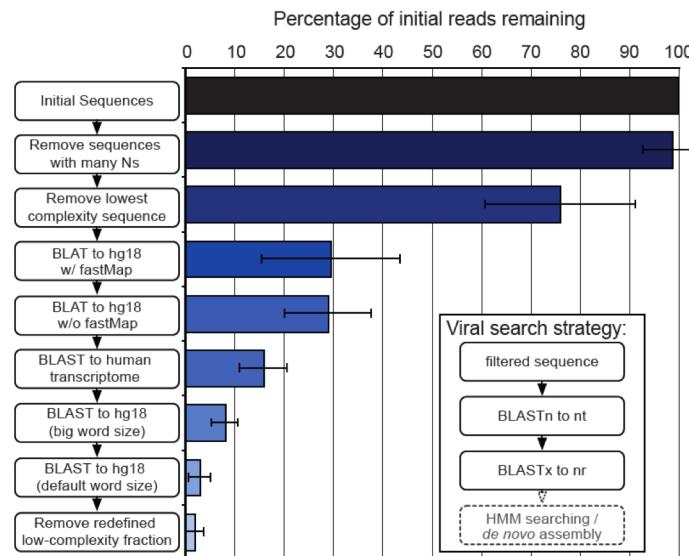
Optimal sequencing depth for RNA-Seq will vary based on the scientific objective of study but here are some general recommendations based on sample type and application:

Sample Type	Reads Needed for Differential Expression (millions)	Reads Needed for Rare Transcript or De Novo Assembly (millions)	Read Length
Small Genomes (i.e. Bacteria / Fungi)	5	30 - 65	50 SR or PE for positional info
Intermediate Genomes (i.e. Drosophila / C. Elegans)	10	70 - 130	50 – 100 SR or PE for positional info
Large Genomes (i.e. Human / Mouse)	15 - 25	100 - 200	>100 SR or PE for positional info

In most cases analysis is more difficult than data generation

You should have a pretty good idea of the analysis strategy before you begin.

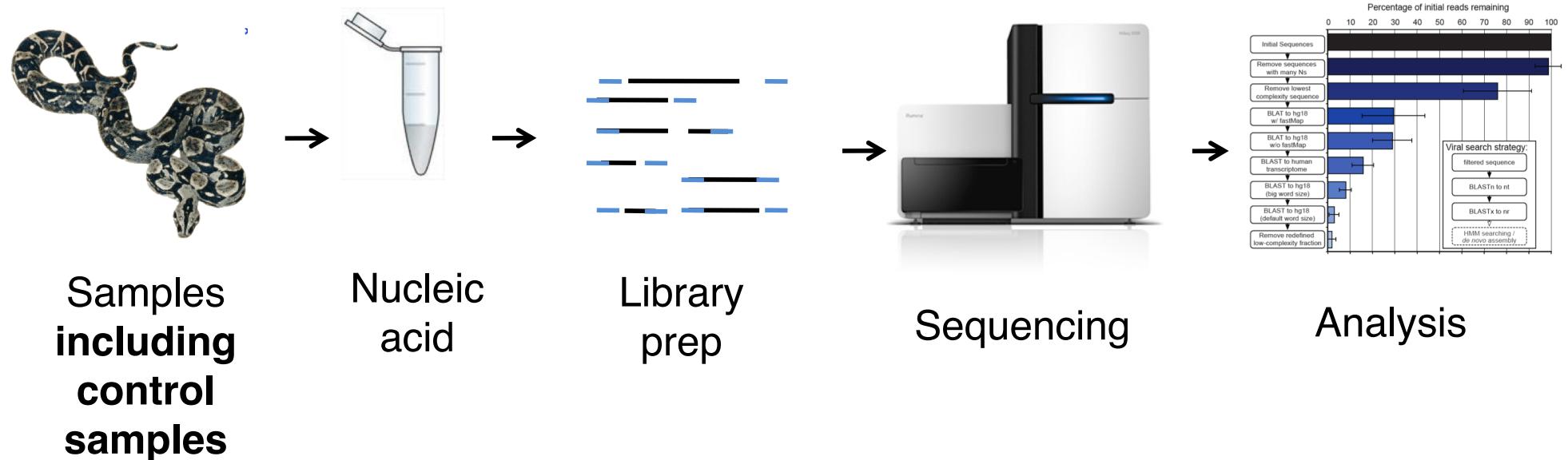
- Is such an analysis feasible? Will it answer your question?
- Are you including the proper controls?
- Do appropriate tools exist? Will you need to design new ones?
- Do you have access to the computational resources necessary?
- If you aren't already comfortable doing the analysis, are you ready to learn how to do it or to work with someone who will do it for you?



Like any experiment, sequencing-based experiments must be properly controlled.

You must include appropriate **positive and negative controls**.

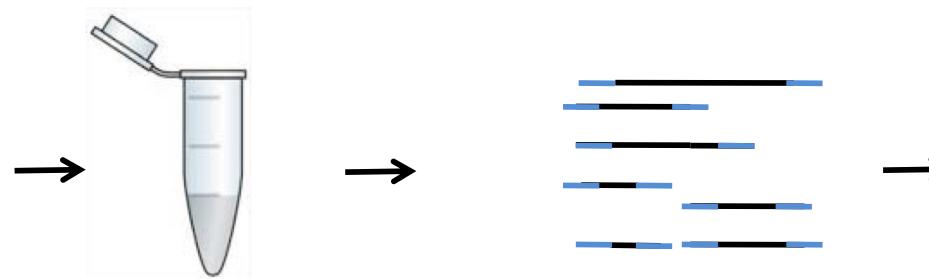
Well designed controls should validate the entire process, from sample processing through library prep and analysis.



NGS-based pathogen discovery



case and control
tissues

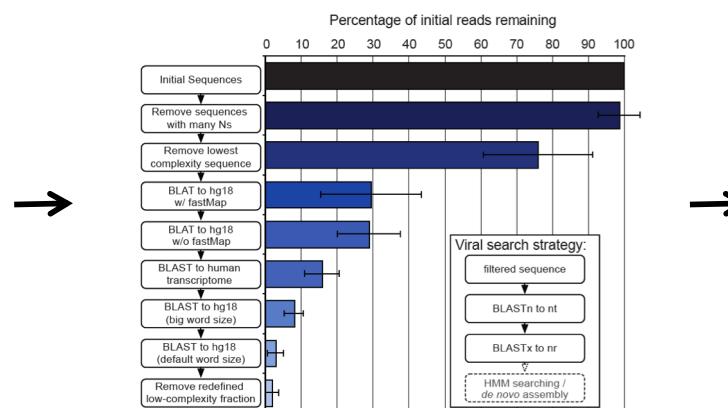


Nucleic acid

Library prep
/ barcode



Illumina
sequencing



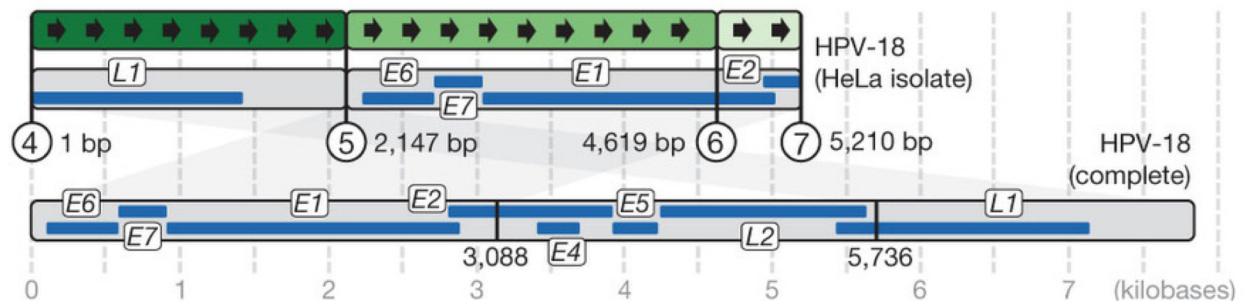
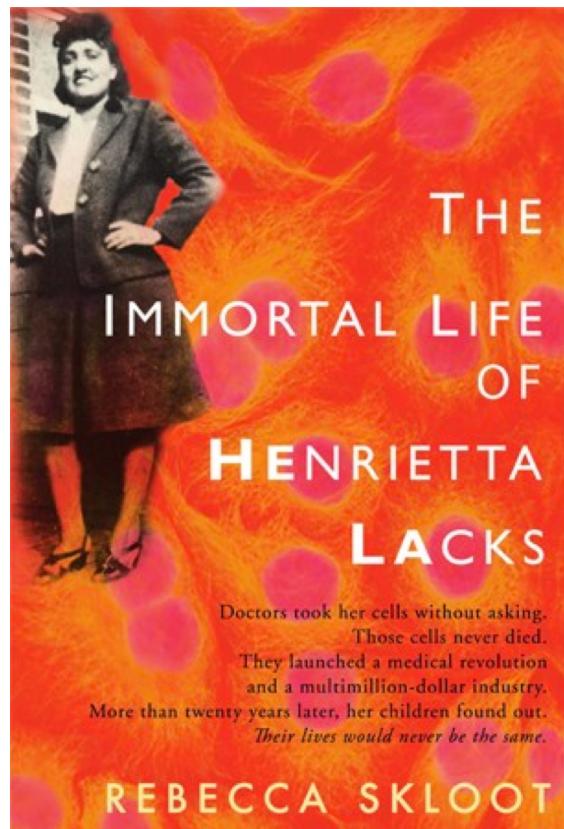
Computational
Analysis



Follow-up

We include positive and negative controls in our pathogen discovery experiments

Positive control: HeLa total RNA



Adey et al (2013) Nature

We include water as a negative control sample in our pathogen discovery experiments



Why water?

Hybrid DNA virus in Chinese patients with seronegative hepatitis discovered by deep sequencing

Baoyan Xu^{a,b,1}, Ning Zhi^{a,1,2}, Gangqing Hu^{c,1}, Zhihong Wan^a, Xiaobin Zheng^d, Xiaohong Liu^a, Susan Wong^a, Sachiko Kajigaya^a, Keji Zhao^{c,3}, Qing Mao^{b,2}, and Neal S. Young^{a,3}

^aHematology Branch and ^cSystems Biology Center, National Heart, Lung, and Blood Institute, Bethesda, MD 20892; ^bInstitute of Infectious Disease, Southwest Hospital, Third Military Medical University, Chongqing 400038, China; ^dDepartment of Embryology, Carnegie Institution for Science, Baltimore, MD 21218

Edited* by Harvey Alter, National Institutes of Health, Bethesda, MD, and approved March 19, 2013 (received for review March 4, 2013)



The Perils of Pathogen Discovery: Origin of a Novel Parvovirus-Like Hybrid Genome Traced to Nucleic Acid Extraction Spin Columns

Samia N. Naccache,^{a,b} Alexander L. Greninger,^{a,b} Deanna Lee,^{a,b} Lark L. Coffey,^c Tung Phan,^c Annie Rein-Weston,^{a,b} Andrew Aronsohn,^d John Hackett, Jr.,^e Eric L. Delwart,^{a,c} Charles Y. Chiu^{a,b,f}

Department of Laboratory Medicine, University of California, San Francisco, California, USA^a; UCSF-Abbott Viral Diagnostics and Discovery Center, San Francisco, California, USA^b; Blood Systems Research Institute, San Francisco, California, USA^c; Center for Liver Disease, University of Chicago Medical Center, Chicago, Illinois, USA^d; Abbott Diagnostics, Abbott Park, Illinois, USA^e; Department of Medicine, Division of Infectious Diseases, University of California, San Francisco, California, USA^f

Reagents can be a source of nucleic acid

TABLE 1 PCR screening of commonly used viral nucleic acid extraction kits for parvovirus-like hybrid virus (PHV-1)^a

Kit	Spin column	PCR result for:							
		Replicase, nt763-1010 (248 nt)		Bridge, nt1554-2044 (491 nt)		Capsid, nt1922-2044 (121 nt)		Capsid + NCR, nt3288-3448 (161 nt)	
		C	F	C	F	C	F	C	F
RNeasy MinElute cleanup kit	RNeasy MinElute column	+	+	-	+	+	+	+	+
	RNeasy minicolumn	+	+	+	+	+	+	+	+
QIAamp UltraSens virus kit	QIAamp minicolumn	+	+	-	+	+	+	+	+
	QIAamp minicolumn	-	+	-	-	+	+	+	+
	QIAamp MinElute column	-	+	-	-	-	+	-	+
PureLink viral RNA/DNA minikit	PureLink viral column	-	-	-	-	-	-	-	-
TRIzol LS kit	NA	-	-	-	-	-	-	-	-
EZ1 viral minikit v2.0	NA	-	-	-	-	-	-	-	-
Water, nuclease-free (Qiagen, Fisher Scientific, and Epicentre)	NA	-	-	-	-	-	-	-	-

^a NCR, noncoding region; C, column elution; F, full extraction; nt, nucleotide; NA, not applicable.

Microbiome studies also impacted

Table 1 List of contaminant genera detected in sequenced negative 'blank' controls

Phylum	List of constituent contaminant genera
Proteobacteria	<p>Alpha-proteobacteria:</p> <p><i>Afipia, Aquabacterium^e, Asticcacaulis, Aurantimonas, Beijerinckia, Bosea, Bradyrhizobium^d, Brevundimonas^c, Caulobacter, Craurococcus, Devosia, Hoeflea^e, Mesorhizobium, Methylobacterium^c, Novosphingobium, Ochrobactrum, Paracoccus, Pedomicrobium, Phyllobacterium^e, Rhizobium^{c,d}, Roseomonas, Sphingobium, Sphingomonas^{c,d,e}, Sphingopyxis</i></p> <p>Beta-proteobacteria:</p> <p><i>Acidovorax^{c,e}, Azoarcus^e, Azospira, Burkholderia^d, Comamonas^c, Cupriavidus^c, Curvibacter, Delftia^e, Duganella^a, Herbaspirillum^{a,c}, Janthinobacterium^e, Kingella, Leptothrix^a, Limnobacter^e, Massilia^c, Methylophilus, Methyloversatilis^e, Oxalobacter, Pelomonas, Polaromonas^e, Ralstonia^{b,c,d,e}, Schlegelella, Sulfuritalea, Undibacterium^e, Variovorax</i></p> <p>Gamma-proteobacteria:</p> <p><i>Acinetobacter^{a,d,c}, Enhydrobacter, Enterobacter, Escherichia^{a,c,d,e}, Nevska^e, Pseudomonas^{b,d,e}, Pseudoxanthomonas, Psychrobacter, Stenotrophomonas^{a,b,c,d,e}, Xanthomonas^b</i></p>
Actinobacteria	<i>Aeromicrobiium, Arthrobacter, Beutenbergia, Brevibacterium, Corynebacterium, Curtobacterium, Dietzia, Geodermatophilus, Janibacter, Kocuria, Microbacterium, Micrococcus, Microlunatus, Patulibacter, Propionibacterium^e, Rhodococcus, Tsukamurella</i>
Firmicutes	<i>Abiotrophia, Bacillus^b, Brevibacillus, Brochothrix, Facklamia, Paenibacillus, Streptococcus</i>
Bacteroidetes	<i>Chryseobacterium, Dyadobacter, Flavobacterium^d, Hydrotalea, Niastella, Olivibacter, Pedobacter, Wautersiella</i>
Deinococcus-Thermus	<i>Deinococcus</i>
Acidobacteria	Predominantly unclassified Acidobacteria Gp2 organisms

The listed genera were all detected in sequenced negative controls that were processed alongside human-derived samples in our laboratories (WTSI, ICL and UB) over a period of four years. A variety of DNA extraction and PCR kits were used over this period, although DNA was primarily extracted using the FastDNA SPIN