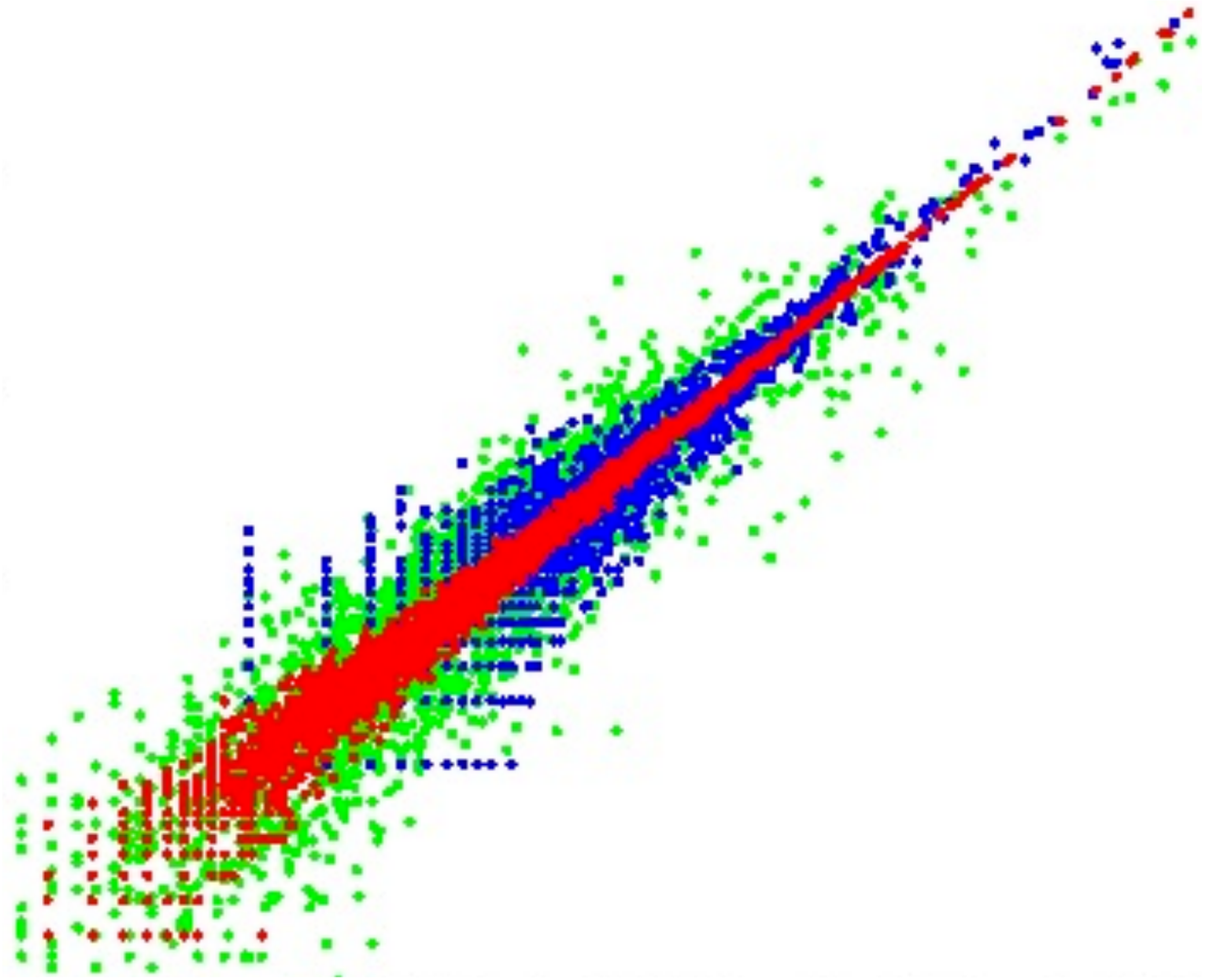
Experimental Design



The datasets might be larger, but basic science principles still apply…

- Controls
- Replication
- Good experimental design

The Importance of Controls in NGS Experiments

PNAS

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]

[a]Hematology Branch and [c]Systems Biology Center, National Heart, Lung, and Blood Institute, Bethesda, MD 20892; [b]Institute of Infectious Disease, Southwest Hospital, Third Military Medical University, Chongqing 400038, China; [d]Department 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)

JVI
Journals.ASM.org

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 as a Source of Contamination

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

		PCR result for:							
		Replicase, nt763-1010 (248 nt)		Bridge, nt1554-2044 (491 nt)		Capsid, nt1922-2044 (121 nt)		Capsid + NCR, nt3288-3448 (161 nt)	
Kit	Spin column	C	F	C	F	C	F	C	F
RNeasy MinElute cleanup kit	RNeasy MinElute column	+	+	−	+	+	+	+	+
RNeasy minikit	RNeasy minicolumn	+	+	+	+	+	+	+	+
QIAamp UltraSens virus kit	QIAamp minicolumn	+	+	−	+	+	+	+	+
QIAamp viral RNA minikit	QIAamp minicolumn	−	+	−	−	+	+	+	+
QIAamp DSP virus kit	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.

Reagents as a Source of Contamination

Contamination also prevalent in 16S microbiome studies

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

Phylum	List of constituent contaminant genera
Proteobacteria	Alpha-proteobacteria:
	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*
	Beta-proteobacteria:
	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*
	Gamma-proteobacteria:
	Acinetobacter[a,d,c], *Enhydrobacter, Enterobacter, Escherichia*[a,c,d,e], *Nevskia*[e], *Pseudomonas*[b,d,e], *Pseudoxanthomonas, Psychrobacter, Stenotrophomonas*[a,b,c,d,e], *Xanthomonas*[b]
Actinobacteria	*Aeromicrobium, Arthrobacter, Beutenbergia, Brevibacterium, Corynebacterium, Curtobacterium, Dietzia, Geodermatophilus, Janibacter, Kocuria, Microbacterium, Micrococcus, Microlunatus, Patulibacter, Propionibacterium*[e], *Rhodococcus, Tsukamurella*
Firmicutes	*Abiotrophia, Bacillus*[b], *Brevibacillus, Brochothrix, Facklamia, Paenibacillus, Streptococcus*
Bacteroidetes	*Chryseobacterium, Dyadobacter, Flavobacterium*[d], *Hydrotalea, Niastella, Olivibacter, Pedobacter, Wautersiella*
Deinococcus-Thermus	*Deinococcus*
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

Salter et al. 2014 *BMC Biology*

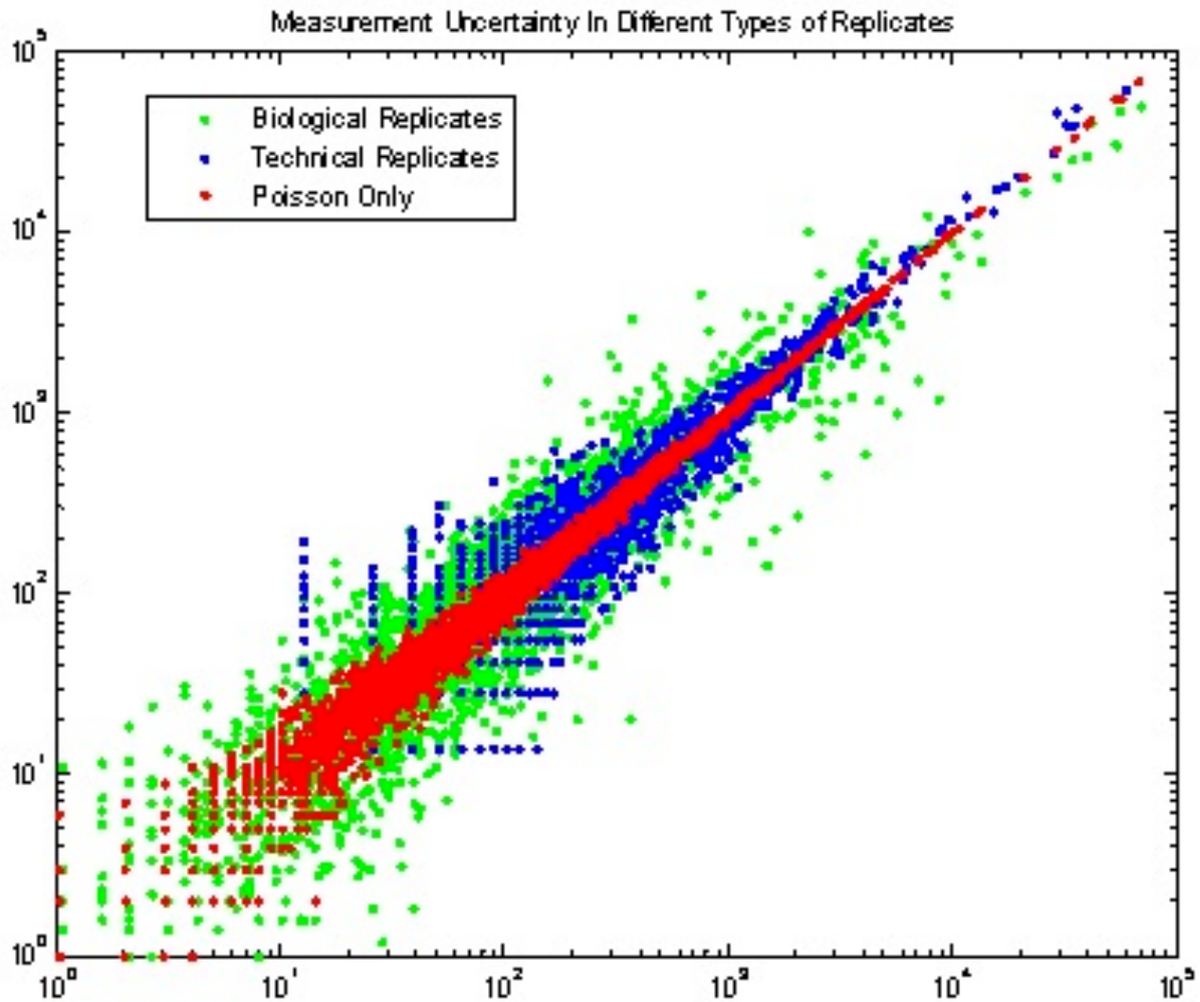
Sources of Variance in NGS Experiments

	Rep1	Rep2	Rep3
Gene A	5	3	12
Gene B	16	25	35
Gene C	10	15	3
Gene D	750	500	500
Gene E	1504	1005	1030

Variation in Read Counts among Replicates

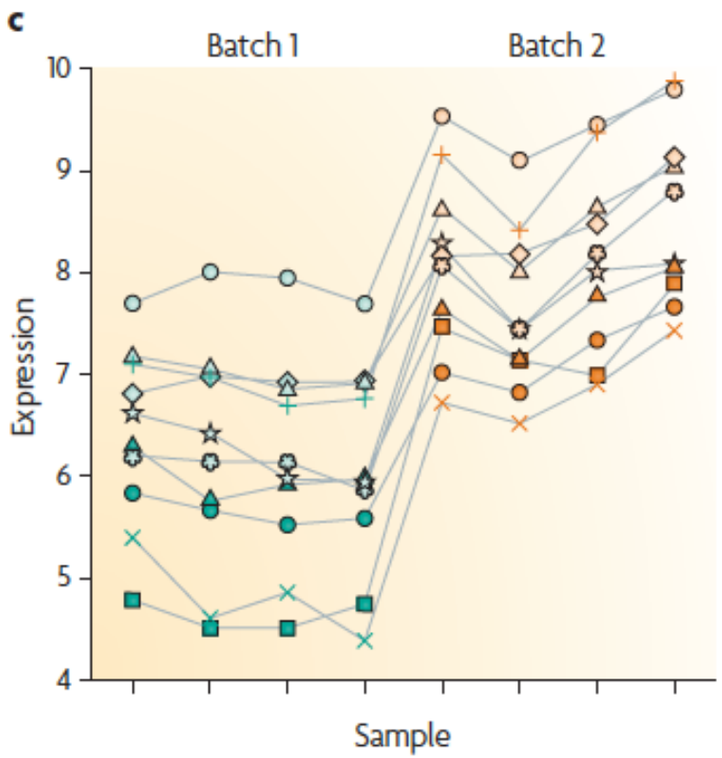
1. Poisson sampling variance
2. Technical variation introduced during library construction and sequence
3. Biological variation between samples

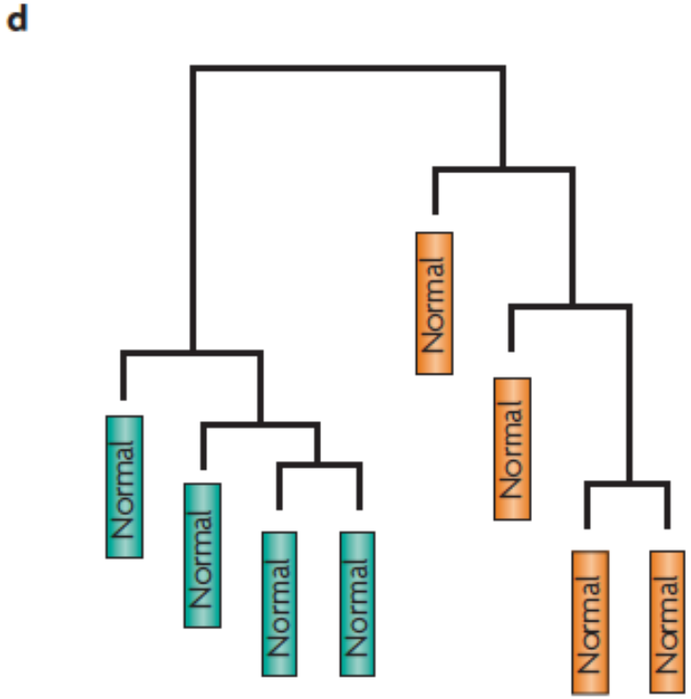
Sources of Variance in NGS Experiments



- Aim for a minimum of three biological replicates in “counting” experiments like differential-expression studies.
- Technical replicates do not address biological variance.
- Doing more replicates is often better than sequencing a small number of replicates more deeply.

Batch Effects and Sources of Bias





- Reagent lots
- Technicians
- Library prep batches
- Sequencer runs/lanes

Leek et al. 2010 *Nat Rev Genet*

Illumina's Recommendations for Reducing Index Hopping

Table 1: Best Practices for Reducing Index Hopping

Mitigation/Recommendation	Benefit/Outcome
Prepare dual indexed libraries with unique indexes[a]	Converts index hopped reads to undetermined
Sequence one 30× human genome per lane[b]	Avoids pooling and index hopping
Remove adapters (cleanup, spin columns, etc)[c]	Reduces levels of index hopping
Store prepared libraries at recommended temperature of –20° C[c]	Reduces levels of index hopping
Pool similar RNA-Seq samples together	Reduces contamination between high and low-expressors

Is this good scientific practice?

Experimental Design Practice

Study 1: Pathogen Discovery

You have observed a die-off of a species of frogs in a local lake and suspect that they may be experiencing an epidemic caused by a novel viral pathogen. You would like to use next generation sequencing to identify candidate viruses that may be responsible for this disease outbreak.

Study 2: Differential Gene Expression

You are interested in how a bacterial infection alters gene expression in a species of shrimp that you study, and you have the ability to experimentally inoculate the shrimp and grow them in culture either with or without the bacterium.

Describe the following features of your experimental design

- Sampling scheme, including plans for replication and/or controls
- Type(s) of nucleic acid to sample and any enrichment/depletion methods
- Sequencing platform and type of sequencing library
- Best practices to be used that will avoid batch effects, pseudoreplication, and artefacts