ARTICLE

BIOTECHNOLOGY and BIOENGINEERING

Exploring the Transcriptome Space of a Recombinant BHK Cell Line Through Next Generation Sequencing

Kathryn C. Johnson,[1] Andrew Yongky,[1] Nandita Vishwanathan,[1] Nitya M. Jacob,[1] Karthik P. Jayapal,[2] Chetan T. Goudar,[2] George Karypis,[3] Wei-Shou Hu[1]

[1]Department of Chemical Engineering and Materials Science, University of Minnesota, 421 Washington Avenue SE, Minneapolis, Minnesota; telephone: +1-612-626-7630; fax: +1-612-626-7246; e-mail: wshu@umn.edu

[2]Global Biologic Development, Bayer HealthCare, Berkeley, California

[3]Department of Computer Science and Engineering, University of Minnesota, Minneapolis, Minnesota

Abstract: Baby Hamster Kidney (BHK) cell lines are used in the production of veterinary vaccines and recombinant proteins. To facilitate transcriptome analysis of BHK cell lines, we embarked on an effort to sequence, assemble, and annotate transcript sequences from a recombinant BHK cell line and Syrian hamster liver and brain. RNA-seq data were supplemented with 6,170 Sanger ESTs from parental and recombinant BHK lines to generate 221,583 contigs. Annotation by homology to other species, primarily mouse, yielded more than 15,000 unique Ensembl mouse gene IDs with high coverage of KEGG canonical pathways. High coverage of enzymes and isoforms was seen for cell metabolism and *N*-glycosylation pathways, areas of highest interest for biopharmaceutical production. With the high sequencing depth in RNA-seq data, we set out to identify single-nucleotide variants in the transcripts. A majority of the high-confidence variants detected in both hamster tissue libraries occurred at a frequency of 50%, indicating their origin as heterozygous germline variants. In contrast, the cell line libraries' variants showed a wide range of occurrence frequency, indicating the presence of a heterogeneous population in cultured cells. The extremely high coverage of transcripts of highly abundant genes in RNA-seq enabled us to identify low-frequency variants. Experimental verification through Sanger sequencing confirmed the presence of two variants in the cDNA of a highly expressed gene in the BHK cell line. Furthermore, we detected seven potential missense mutations in the genes of the growth signaling pathways that may have arisen during the cell line derivation process. The development and characterization of a BHK reference transcriptome will facilitate future efforts to understand, monitor, and manipulate BHK cells. Our study on sequencing variants is crucial for improved understanding of the errors inherent in high-throughput sequencing and to increase the accuracy of variant calling in BHK or other systems.

Biotechnol. Bioeng. 2014;111: 770–781.

KEYWORDS: BHK; Baby Hamster Kidney cells; Syrian hamster; transcriptome; RNA-seq; sequence variants

Introduction

Baby hamster kidney (BHK) cell lines, which originated from primary cultures of neonatal Syrian hamster kidney tissue in the 1960s (Macpherson and Stoker, 1962; Stoker and Macpherson, 1964), are widely used in biomedical research. These cell lines are also used in the industrial production of recombinant therapeutic proteins (Dingermann, 2008; Jiang et al., 2002) as well as veterinary vaccines (Pay et al., 1985; Radlett et al., 1985). The Syrian hamster is also an important animal model in the research of infectious disease (Paessler et al., 2004; Requena et al., 2000; Xiao et al., 2001), cardiomyopathy (Crespo and Escobales, 2008), diabetes (Bhathena et al., 2011; Popov et al., 2003), atherosclerosis (Dillard et al., 2010; Jové et al., 2012), and neural plasticity (Staffend and Meisel, 2012).

Rapid advances in DNA sequencing technologies should now permit the exploitation of the BHK transcriptome, genome, and epigenome to facilitate cell line and process development. To embark on such an effort one needs high-quality, annotated reference sequences. GenBank contains approximately 3,000 nucleotide records for BHK or Syrian hamster that vary in their state of annotation or sequence completeness. An effort to obtain a Syrian hamster genome is

Kathryn C. Johnson and Andrew Yongky contributed equally to this work.

Correspondence to: W.-S. Hu

Contract grant sponsor: Bayer HealthCare
Contract grant sponsor: NSF Graduate Research Fellowship
Received 5 June 2013; Revision received 5 September 2013; Accepted 7 October 2013
Accepted manuscript online 12 October 2013;
Article first published online 19 November 2013 in Wiley Online Library (http://onlinelibrary.wiley.com/doi/10.1002/bit.25135/abstract).
DOI 10.1002/bit.25135

presently underway at the Broad Institute (NCBI Bioproject 77669). We developed an annotated BHK transcriptome by assembling RNA-seq data from a recombinant BHK (rBHK) cell line and Syrian hamster tissues and annotating the resulting contigs against homologous sequences. Having obtained high coverage of many canonical pathways, we also gained valuable insights into some characteristics of BHK gene expression, including elements of energy metabolism and glycosylation.

With growing recognition of the contribution of single nucleotide variants to genetic diversity, high-throughput sequencing has become a preferred method for large-scale discovery of such variants. The 1,000 Genomes Project has employed this approach to identify single nucleotide polymorphism (SNP) sites in the human genome, an effort that resulted in an immense catalog of more than 14.4 million and 12,758 SNPs in the human genome and exome, respectively (Abecasis et al., 2010). The major challenge in utilizing high-throughput sequencing for variant detection lies in distinguishing the real variants from errors introduced during the sequencing process. Several methods have been developed to identify variants from whole genome sequencing data (DePristo et al., 2011; Lam et al., 2012; Reumers et al., 2012); however, performing such analyses in cell lines such as BHK contributes additional complications.

Whereas most nucleotide variant analyses have been performed on DNA from diploid individuals, cell lines have a propensity to accumulate mutations and copy number differences over the course of continual culture, meaning that the sample being sequenced represents a heterogeneous population. The fraction of the population harboring such mutations may remain low unless it offers some growth advantage or recurs repeatedly, but the variant frequency will rarely reach the level of heterozygous variants that most variant analysis methods are designed to detect. A majority of such mutations in recombinant cell lines went unnoticed until recently because few cell lines have been sequenced to adequate depth to detect such low-frequency mutations. Even with high-throughput genome sequencing, the depth of coverage employed is rarely more than 100. RNA-seq, on the other hand, sequences the transcripts of abundant genes to much greater depth, sometimes to tens of thousands of reads per base (Jacob et al., 2010). However, the development of methods to identify variants in whole transcriptome sequencing data is an area that has not been explored as well as its whole genome counterpart. The very high coverage obtained for BHK transcripts enabled us to investigate sequence variants in this cultured cell line and to develop a workflow for variant detection in RNA-seq data.

Materials and Methods

Cell Culture, Tissue Collection, and RNA Preparation

All cell cultures were performed at Bayer. A BHK-21 derived recombinant protein producing cell line (rBHK) was cultured in animal component-free and chemically defined medium at 37°C under 5% CO_2 in a 50 mL volume in a 250-mL shake flask at 110 rpm. Cells were harvested in exponential growth phase, lysed in 1 mL TRIzol reagent (Invitrogen, Carlsbad, CA), and stored at −80°C. Brain and liver tissues from an adult female Syrian hamster were kindly provided by Dr. Robert Meisel snap-frozen in liquid nitrogen. Approximately 25 μg of frozen tissue was homogenized in 1 mL TRIzol before proceeding with RNA extraction. Total RNA was extracted according to the manufacturer's protocol.

Illumina Library Preparation and Sequencing

Illumina sequencing was performed by the University of Minnesota's Biomedical Genomics Center (St. Paul, MN), BGI (Shenzhen, China), and the National Center for Genome Resources (Santa Fe, NM). A single sample from the rBHK cell line was used to construct two separate libraries (rBHK1, rBHK2) which were sequenced independently. Barcoded brain and liver samples were sequenced together in a single lane. RNA-seq library preparation was performed according to the Illumina TruSeq protocol. All samples were sequenced on the Illumina HiSeq 2000 as short-insert paired-end libraries with read lengths of 90 bp (rBHK2) or 100 bp (rBHK1, brain, liver). Raw fastq files were processed to remove sequencing adapters and to trim low-quality bases (Phred quality score <15) from both ends of the reads. Reads trimmed to <45 bp were discarded. Details on sequencing assembly and annotation are in Supplemental Methods.

Alignment and Transcript Level Quantification

All pre-processed reads were mapped to the final assembly as single-end reads for the purposes of estimating expression levels, using the gap-enabled aligner BWA v.0.5.9 (Li and Durbin, 2009) with a maximum of three mismatches allowed. The SAMtools suite of algorithms (v.0.1.18) (Li et al., 2009) was used to process the resulting alignments. Paired-end mapping was also performed to assess the proportion of properly paired reads. For local realignment and variant filtering, see Supplemental Methods. Uniquely mapped reads were used to estimate transcript levels. To normalize mapped read counts by contig length and library depth, values were expressed in units of RPKM (reads per kilobase per million reads mapped) (Mortazavi et al., 2008). When multiple contigs represented the same gene, RPKM was calculated by dividing the total number of reads mapping to those contigs by the sum of their lengths.

For cross-library comparisons of transcript levels, we normalized the mapped read counts across libraries using the upper quartile normalization method (Bullard et al., 2010). When multiple contigs represented the same gene, the contig with the maximum normalized number of reads mapped across all libraries was chosen as the contig to represent that gene. Sequencing depth for each gene was calculated as upper-quartile normalized read counts for the gene divided by the contig length (reads/kb).

Sequence Verification by Sanger Sequencing

RNA from rBHK was reverse-transcribed using either the SuperScript III Reverse Transcriptase (SSIII-RT) (Invitrogen) or the Moloney Murine Leukemia Virus Reverse Transcriptase (M-MuLV RT) (New England BiolabsPlease provide the manufacturer location: city, state (if USA) and country name.) with gene-specific primers. cDNA was amplified with gene-specific primers targeting the candidate variant-containing regions containing using Phusion High-Fidelity DNA polymerase (New England Biolabs). PCR products were gel-purified using the QIAquick Gel Extraction kit (Qiagen Please provide the manufacturer location: city, state (if USA) and country name.), cloned into the TOPO-vector (Invitrogen) and transformed into One Shot TOP10 *Escherichia coli. E. coli* cells were spread on agar plates under 50 μg/mL kanamycin selection and colonies were individually picked and expanded in LB broth containing 50 μg/mL kanamycin. Plasmids were purified using the Qiaprep Spin Miniprep kit (Qiagen) and subjected to Sanger sequencing.

Results

Transcriptome Assembly

A total of 40.5 Gbp of paired-end sequence was obtained from the BHK cell line and Syrian hamster tissues (Supplementary Table SI). After data pre-processing, 385 million reads were used for assembly. Using the Oases assembler, multiple assemblies of the rBHK libraries were performed to optimize the k-mer value (k) as described in the Supplemental Methods. Among those values tested (57, 59, 61, 63), the $k = 63$ assembly yielded the fewest contigs (157,045) and the greatest N50 length (668 bp). All values of k gave a similar percentage of reads used (∼70%). These assemblies were merged using CD-HIT-EST and Phrap. The brain, liver, and unassembled rBHK reads were separately assembled ($k = 63$) and the resulting contigs were then merged with CD-HIT-EST and Phrap. At the Phrap step 6,170 Sanger ESTs previously obtained from rBHK and parental BHK were also included (Yee, 2008).

The final 221,583 contigs have an average length of 577 bp and a maximum length of 24,447 bp (Table I). The length distribution of all contigs is provided in Supplementary Figure S1. Sixty-eight to 83% of the reads from each library mapped back to the assembly using single-end alignment (Supplementary Table SII). Eighty-nine to 95% of paired reads in each library mapped back to the assembly with a separation distance consistent with the insert size. A total of 6,053 (98.1%) of the Sanger ESTs were incorporated into contigs with Illumina reads. The 110 unassembled ESTs longer than 200 bp were retained in the assembly.

Table I. Final assembly statistics.

Number of contigs	221,580
Average contig length (bp)	577
Median contig length (bp)	336
N50 contig length (bp)	758
Largest contig length (bp)	24,447
Total transcriptome length (bp)	127,759,291

Annotation and Pathway Coverage

A total of 129,722 contigs were annotated against one of the seven databases queried (Supplemental Methods). 65,426 contigs were assigned an Ensembl mouse gene ID; these IDs represented 15,145 unique genes. The well-annotated genes were represented, on average, by three contigs. Highly fragmented sequences were mostly from genes with long transcripts as indicated by their mouse homologs.

We assessed the coverage of 186 canonical pathways in KEGG by the annotated contigs. Supplementary Figure S2 depicts several of these pathways. On average, 86% of genes in each of the 186 pathways were represented in the assembly, and only five pathways were less than 50% represented. In metabolic gene sets, many missing genes are isozymes that exhibit tissue-restricted expression. In pyruvate metabolism, for example, the only genes not detected are the testes-specific isozymes Pdha2 and Ldhc; complementary isozymes (Pdha1, Pdhb; Ldha, Ldhb) are represented.

This assembly captures a substantial portion of the Syrian hamster transcript sequence in GenBank. Blastn alignment to 2723 GenBank entries for Syrian hamster yielded hits for 70% of those sequences (E-value $\leq$ 1E−04). Of the remaining GenBank entries, many come from patents, represent genomic DNA, and/or represent genes that exhibit tissue-restricted expression.

Homology

We compared the homology of all our contigs at the nucleotide level to Chinese hamster and mouse. The distributions of the percentage identities are shown as histograms in Figure 1a. The median nucleotide identity among the sequences compared for the two hamster species was 94%, while that for Syrian hamster and mouse was 91%. Substantially larger fractions of contigs had a higher identity with Chinese hamster (Fig. 1b). The GC content of the Syrian hamster transcriptome was also closest to that of the Chinese hamster (Supplementary Table SIII).

RNA-Seq Analysis

Range of Gene Expression

Transcript expression levels were estimated for each library (see Materials and Methods Section). Figure 2 depicts the distributions of transcript expression levels for the rBHK1 and liver libraries; those for rBHK2 and brain appear similar, respectively (data not shown). In the rBHK libraries, the exogenous dihydrofolate reductase (Dhfr) gene was among the 20 most abundant genes and the recombinant product gene was among the top 2% of genes, while in the brain and liver libraries many of the most abundant genes

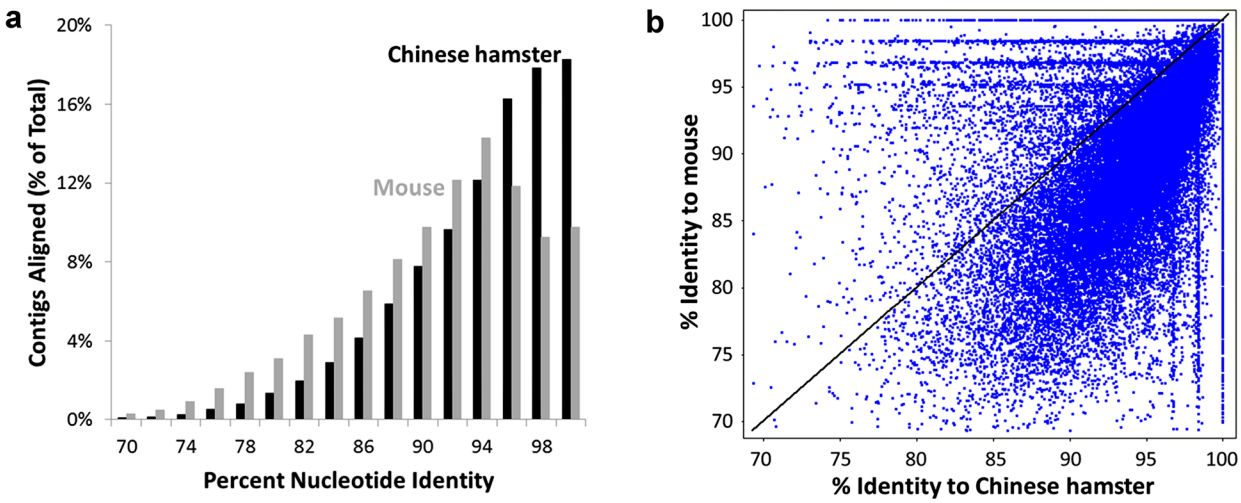

Figure 1. **a**: Frequency distribution of BHK contigs nucleotide identities with Chinese hamster (black bars) and mouse (gray bars). Values are given as a percentage of contigs aligned because of a difference in the number of BHK contigs aligning to each database. **b**: Scatter plot of BHK contigs percentage nucleotide identities with Chinese hamster versus their percentage identities with mouse. Contigs shown are only those with BLAST alignments to both Chinese hamster ESTs and Ensembl mouse transcripts.

were tissue-specific,(e.g., myelin basic protein in brain, albumin and transferrin in liver). In all four libraries, peroxiredoxin, mitochondrial cytochrome *c* oxidase III, and mitochondrial NADH dehydrogenase 1 were consistently among the 20 highest expressed genes, and other housekeeping genes (Gapdh, β-actin) were also expressed at consistently high levels. The median values were similar across all libraries (4–7 RPKM). The longer left-hand tails of the rBHK distributions versus those of the tissues can be attributed to the higher depth of sequencing employed for the cell line. Because detection of the rarest transcripts by sequencing can be a stochastic process, interpretation of very low RPKM values should be performed with caution. However, with our depth of sequencing, 0.1 RPKM corresponded to an average of 1× coverage. The data thus show that even for contigs expressed at low levels (<1 RPKM), there was a very wide dynamic range. Transcript expression levels in *N*-glycan biosynthesis and glycolysis, both of which strongly influence product titer and quality, are discussed below.

N-Glycan Biosynthesis

All genes in the KEGG *N*-glycan biosynthesis pathway are represented in the assembly. Figure 3 depicts the steps of *N*-glycan processing and provides relative expression levels of these genes for rBHK1 and liver expressed as upper-quartile normalized read counts. The data from the two rBHK libraries are similar across a range of expression levels (Supplementary Fig. S3). Subunits of the oligosaccharyltransferase complex (Rpn1, Rpn2, Dad1, Ddost), which

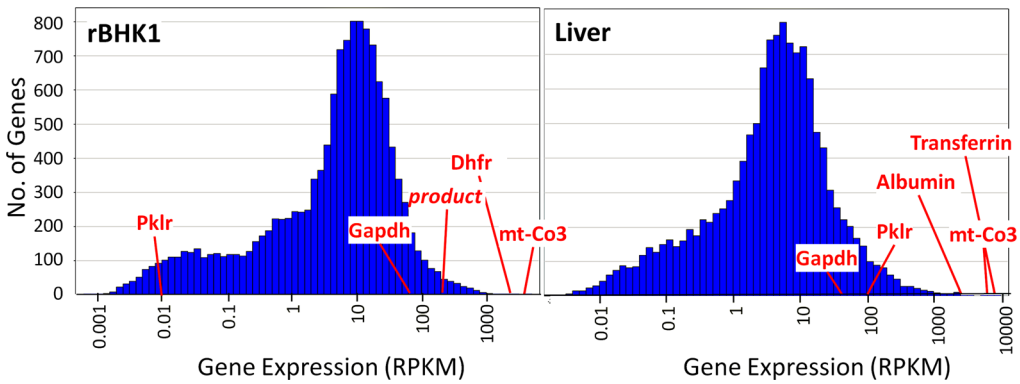

Figure 2. Frequency distributions of RPKM values for (**a**) rBHK1 and (**b**) Syrian hamster liver.

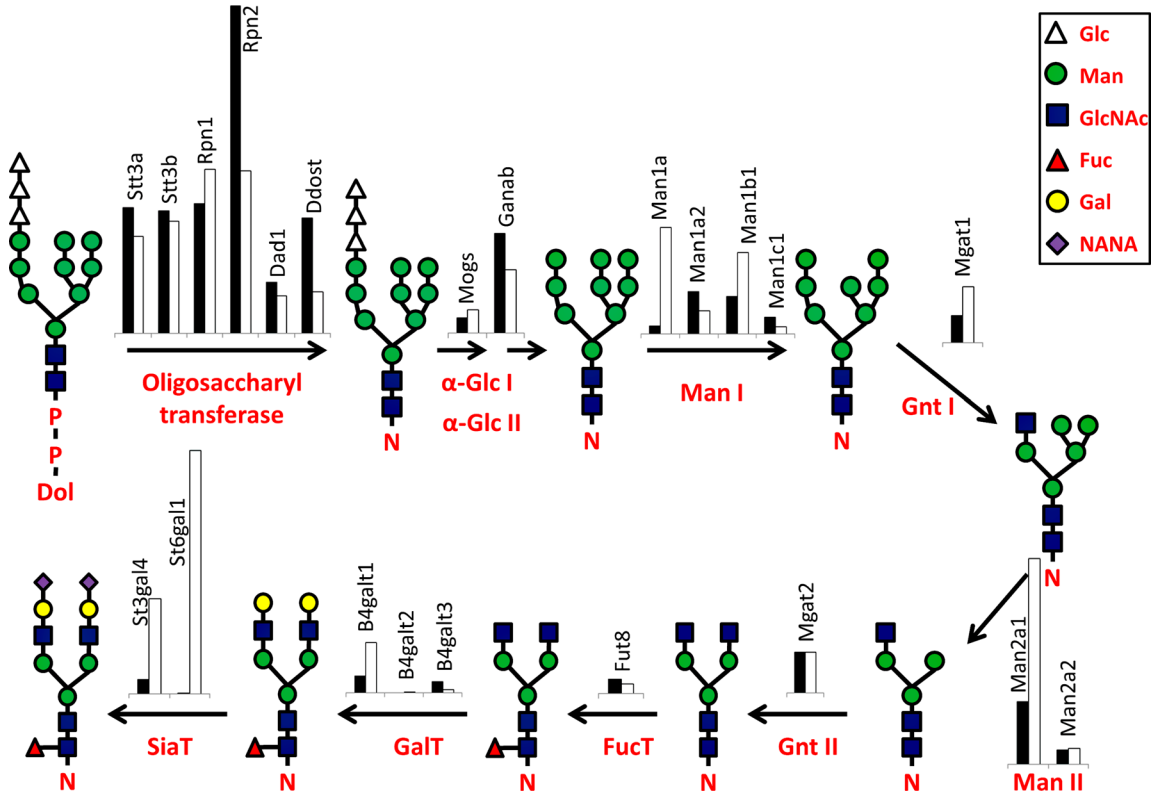

Figure 3. *N*-glycan biosynthesis pathway gene expression levels. Upper-quartile normalized read counts are shown for selected genes in the *N*-glycan biosynthesis pathway. Black bars (rBHK1) represent the rBHK cell line, and white bars represent Syrian hamster liver. All bar charts are on the same *y*-axis scale. OST, oligosaccharide transferase; α-Glc I, α-glucosidase I; α-Glc II, α-glucosidase II; Man I, mannosidase I; Gnt I, GlcNAc transferase I; Man II, mannosidase II; GnT II, GlcNAc transferase II; FucT, fucosyltransferase; GalT, galactosyl transferase; SiaT, sialyltransferase.

catalyzes the transfer of a pre-assembled oligosaccharide from a lipid anchor to an asparagine residue on a recipient protein, were among the most highly expressed genes in this pathway for both liver and rBHK. These subunits were each expressed at ~3,700–23,000 reads/kb in rBHK; in contrast, the median expression levels of all genes in this pathway was ~1,100 reads/kb for rBHK, and the median expression level across all genes for this library was 470 reads/kb. Except for some apparent variability in each cell type's isozyme preferences (ex. mannosidase I), most of this pathway's genes appeared to be expressed at similar levels in rBHK and liver.

Sialic acid moieties are often added to a galactose residue by a sialyltransferase to become the terminal residues on *N*-glycans. While human and mouse cell lines express both α-2,3- and α -2,6-sialyltransferases, hamster cell lines express primarily α-2,3-sialyltransferases (Butler, 2006). The 2,3-transferase St3gal4 was expressed in the rBHK cell line, while St6gal1, a 2,6-transferase, was nearly absent. This 2,6-transferase, however, was expressed at a moderately high level in liver.

The GlcNAc transferases GnTIII, GntIV, and GntV were also assembled. GnTIII, which adds a bisecting GlcNAc, is virtually absent in hamster cell lines (Butler, 2006), and was expressed at a low level in both liver (7 reads/kb) and the cell line (12 reads/kb); it was however more highly expressed in brain (3,300 reads/kb). GntIV and GntV, which create tri- and tetra-antennary branched structures, were detected at moderate levels in all libraries.

Glycolysis

Fifty-three of the 58 genes belonging to the KEGG glycolysis gene set are represented among the contigs. Although their median normalized read count ranged from 690 to 2,800 reads/kb among the four libraries, this pathway contained some of the most abundant genes in the entire transcriptome. In each library 9–15 genes were represented by more than 10,000 reads/kb.

Many steps of glycolysis can be catalyzed by multiple isozymes. The isozymes' relative expression levels, which can potentially affect the cells' metabolism, shifted significantly between cell types (Fig. 4). In rBHK, hexokinase 1 (Hk1) was the dominant isoform of hexokinase, while Hk2 was still present and Hk3 was nearly absent (Fig. 4a). The high level of Gck in liver versus brain recapitulated this isozyme's known

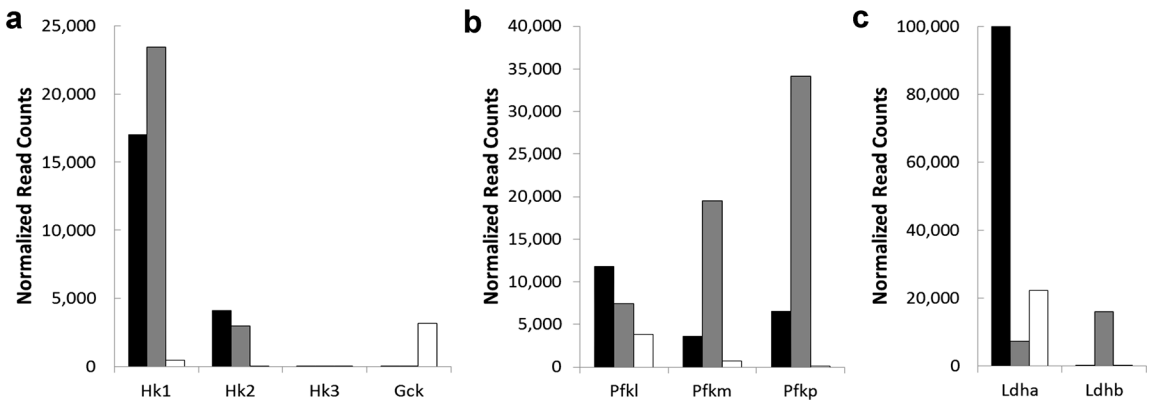

Figure 4. Expression levels of selected glycolytic enzymes. Black bars = rBHK1, gray bars = Syrian hamster brain, white bars = Syrian hamster liver. **a**: Isozymes of hexokinase; (**b**) isozymes of phosphofructokinase; (**c**) isozymes of lactate dehydrogenase (no contig assembled for Ldhc).

tissue specificity. Among the three isoforms of phosphofructokinase, Pfkp was essentially absent in liver, and both brain and rBHK expressed all three isozymes (Fig. 4b). Lactate dehydrogenase (Ldh) isozymes are expressed at exceptionally high levels (Fig. 4c). No contig was assembled for Ldhc, an isozyme known to be primarily expressed in testis. Of the other two Ldh isozymes, Ldhb is known to be preferentially expressed in brain, consistent with our results. This transcript was essentially absent in the cell line (4–26 reads/kb) relative to Ldha. Ldha was extremely abundant in the cell line; in fact, it was one of the most highly expressed genes in the rBHK libraries (~100,000 reads/kb).

Large-Magnitude Differences Between Tissues and Cell Line

Two thousand three hundred sixty-four genes demonstrated very large-magnitude (>10×) fold-changes between the rBHK cell line and both hamster tissues (see Supplemental Methods). The functional classes that contain a large number of these genes include cell cycle and p53 signaling (increased expression in rBHK relative to both tissues) as well as cell adhesion molecules and G-protein-coupled receptors (decreased expression in rBHK). The large differences in the cell cycle functional class (Supplementary Fig. S4) are not surprising given the fast cell growth seen in the cell line versus the relatively quiescent nature of the tissues. cMyc is also 17-fold higher in the cell line than in liver and 60-fold higher than in brain. cMyc is an oncogene frequently up-regulated in cancers whose role in the regulation of proliferation, cell growth, and other functions has been studied extensively [reviewed by (Meyer and Penn, 2008)]. Decreased expression of cell adhesion molecules likely reflects the fact that the cell line has been adapted to growth in suspension. These molecules include integrins, cadherins, and junction adhesion molecules, among others. Other strongly down-regulated genes include glutamine synthetase (Glul), α2,6-sialyltransferase (St6gal1), lactate dehydrogenase B (Ldhb), and glycerol-3-phosphate dehydrogenase (Gpd1).

Sequence Variant Analysis

Identification of Single Nucleotide Variation

BHK cells have been propagated in culture for perhaps thousands of generations. A large number of mutations are expected to have accumulated in the chromosomes. In this study, we attempted to identify single nucleotide variations in the transcripts of BHK using RNA-seq. Most algorithms for single nucleotide variant detection such as GATK (McKenna et al., 2010) or SAMtools were developed for cases where DNA sequencing was performed for individuals with diploid chromosomes. Single nucleotide variations in the DNA in each of these individuals would result in allele frequencies of either 0.5 or 1 representing heterozygous or homozygous variants, respectively. However, in BHK cells, chromosomal copy number change may have occurred due to long-term propagation in culture such that the cells are no longer diploid. Furthermore, different subpopulations of the cells may have acquired different types of mutations, resulting in a heterogenous population. Thus, identification of single nucleotide variations in the transcriptome of the BHK cells using these algorithms may not be appropriate.

We devised a workflow to detect single nucleotide variants in RNA-seq reads (see Supplemental Methods for details; also see Supplementary Figs. S5 and S9–15). An initial variant scan was made using VarScan v. 2.2.10 (Koboldt et al., 2012), which employs the following heuristics: a minimum coverage of 8 reads at the variant position, a minimum of three reads supporting the variant base (henceforth referred to as variant reads), and a minimum variant frequency of 1% of the total read depth at that position. The base quality score and mapping quality score were set to minimum values of

30 and 20, respectively. The Illumina protocol sequenced both strands of the cDNA, therefore true variants should be represented by variant reads in both directions. Variants that are supported by reads in only a single read direction were filtered out.

The performance of the algorithms was first assessed using the liver and brain libraries. Initially, variants were only called from contigs of more abundant transcripts (RPKM brain $\geq$8, RPKM liver $\geq$8, 1/2 $\leq$ RPKM brain/RPKM liver $\leq$2). 2,538 and 3,070 potential variants were identified in the liver (library A) and brain (library B), respectively, with 1,416 in common (denoted by $A_1 \cap B_1$ in Fig. 5 and Table II), while 1,122 and 1,654 were exclusive to liver ($A_1 \cap (\neg B_1)$) and brain ($(\neg A_1) \cap B_1$), respectively. The variants exclusive to either liver or brain occur mostly at low frequency, and potentially originated from random sequencing error, which has been estimated to occur at a rate of ~0.1–1% (Glenn, 2011).

To distinguish potentially true variants from random errors, a Poisson model was applied to each library independently. In some cases a variant that was initially found to be common in both libraries is now retained in only one. The resulting variants were classified into five categories (Fig. 5): (1) $\{A_2 \cap (A_1 \cap B_1)\}$, variants retained in library 1 which were shared in both libraries before filtering with a Poisson model, (2) $\{B_2 \cap (A_1 \cap B_1)\}$, variants retained in library 2 which were shared in both libraries before filtering, (3) $\{A_2 \cap B_2\}$, variants common to the two libraries both before and after filtering, (4)$\{A_2 \cap (\neg B_1)\}$, variants exclusive to library 1 before and after filtering, and (5) $\{B_2 \cap (\neg A_1)\}$, variants exclusive to library 2 before and after filtering.

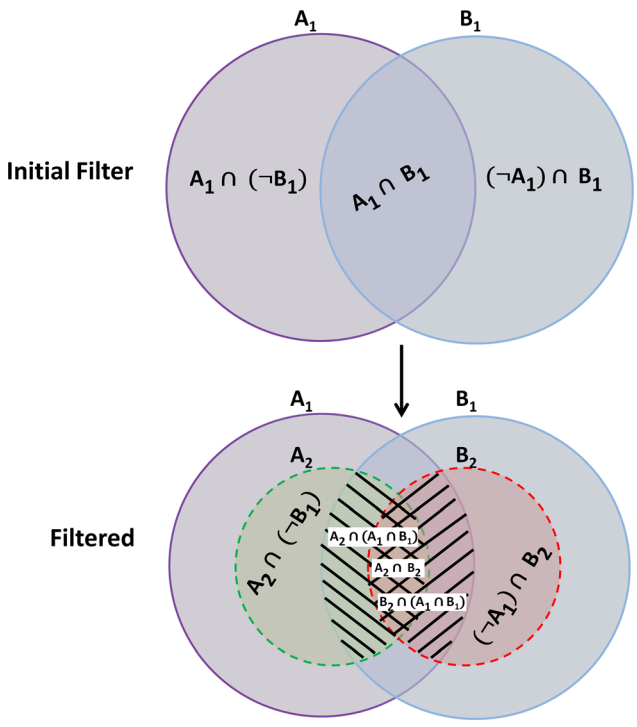

Figure 5. Categories of variant at initial call and after filter.

Upon application of the Poisson filter, we obtained 1,327 high-confidence variants in liver and 1,466 in brain (Table IIa), with 1,071 in category (1), 1,115 in (2), 946 in (3), 256 in (4), and 351 in (5). Comparison of the number of variants in each category before and after all filters shows that those in categories (1)–(3) are largely retained after filtering (76%, 79%, and 69%, respectively), while few in categories (4) and (5) remain (23% and 21%, respectively). The result indicates reasonable selectivity of the Poisson filter (see also Supplementary Fig. S6a). A majority of the low-frequency variants were removed while variants that are likely to be heterozygous germline SNPs were retained (Fig. 6a and b). Although a considerable number of variants are retained in categories (4) and (5), they still likely represent high-confidence variants. Inspection of variants in categories (4) and (5) indicates that they are mostly located in genes that are specifically expressed in the corresponding tissues (data not shown); for example, variants in the liver-specific gene Gstm7 are retained in category (4) and variants in the brain-specific gene Tubb2c are retained in category (5).

Further, the Fisher Exact test (FET) was used to eliminate potential strand bias in the detected variants. While systematic error causes a variant to be present in *only* one direction of the reads, a combination of systematic error and random base-calling error may cause a variant to appear in *both* strands but heavily biased towards one of the two. The FET compares the read distribution in both directions of the variant reads to the distribution in both directions of the reference reads (see Supplementary Materials for details). The application of the FET resulted in a meager improvement in the variants called (Table IIa, last row).

The same analyses were performed by comparing the variants found in rBHK1 (library A) and rBHK2 (library B). These two libraries have higher coverage than the tissue samples. Therefore, we employed a high abundance level cutoff (RPKM rBHK1 $\geq$ 32, RPKM rBHK2 $\geq$ 32, 1/2 $\leq$ RPKM rBHK2/RPKM rBHK1 $\leq$ 2). The initial analysis resulted in 6,957 and 11,293 potential variants called in the rBHK1and rBHK2 libraries, respectively, with 5,080 variants in common (denoted by $A_1 \cap B_1$ in Table IIb), while 1,877 were exclusive to rBHK1 ($A_1 \cap (\neg B_1)$) and 6,213 were exclusive to rBHK2 ($(\neg A_1) \cap B_1$). Upon application of the Poisson and FET filters, we obtained 3,544 high-confidence variants in rBHK1 and 5,326 in rBHK2 (Table IIb), with 3,149 in category (1), 3,797 in (2), 3,260 in (3), 514 in (4), and 1,529 in (5). Unlike the variants in the tissues, most variants in the cell lines are of low frequencies (<10%) (Fig. 6c and d). It is very likely that they were not detected in the other library due to their low frequencies.

After demonstrating the filters' performances in the high RPKM contigs, the analyses were performed for the entire collection of contigs. From the tissue libraries, 6,608 high-confidence variants were obtained in the liver and 8,621 in the brain after Poisson and FET filters (Supplementary

Table II. Summary of potential variants in contigs with high RPKM upon application of the Poisson filtering criteria.Please check the presentation of this table.

Filter criteria			$A_1 \cap B_1$		$A_1 \cap (\neg B_1)$	$(\neg A_1) \cap B_1$
(a) In the liver (A) and the brain (B)						
(I)	Initial call		1,416 (100%)		1,122 (100%)	1,654 (100%)
		$\mathbf{A_2 \cap (A_1 \cap B_1)}$	$\mathbf{B_2 \cap (A_1 \cap B_1)}$	$\mathbf{A_2 \cap B_2}$	$\mathbf{A_2 \cap (\neg B_1)}$	$\mathbf{B_2 \cap (\neg A_1)}$
(II)	(I) + Poisson	1,071 (75.6%)	1,115 (78.7%)	978 (69.1%)	256 (22.8%)	351 (21.2%)
(III)	(II) + FET	1,060 (74.8%)	1,102 (77.8%)	963 (68%)	251 (22.4%)	341 (20.6%)
(b) In rBHK1 (A) and rBHK2 (B)						
(I)	Initial call		5,080 (100%)		1,877 (100%)	6,213 (100%)
		$\mathbf{A_2 \cap (A_1 \cap B_1)}$	$\mathbf{B_2 \cap (A_1 \cap B_1)}$	$\mathbf{A_2 \cap B_2}$	$\mathbf{A_2 \cap (\neg B_1)}$	$\mathbf{B_2 \cap (\neg A_1)}$
(II)	(I) + Poisson	3,838 (75.5%)	4,055 (79.8%)	3,537 (69.6%)	514 (27.3%)	1,844 (29.7%)
(III)	(II) + FET	3,149 (62.0%)	3,797 (74.7%)	3,260 (64.2%)	395 (21.0%)	1,529 (24.6%)

Table SIVa, last row). The final variants were distributed as follows: 4,733 in category 1, 5,124 in category 2, 4,215 in category 3, 1,875 in category 4 and 3,497 in category 5. Similar analyses on cell line libraries resulted in 14,375 high-confidence variants in rBHK1 and 21,165 in rBHK2 (Supplementary Table SIVb, last row). The final variants were distributed as follows: 12,035 in category 1, 12,685 in category 2, 11,878 in category 3, 2,340 in category 4, and 8,480 in category 5. A considerable number of variants were still retained in categories 4 and 5, consistent with previous findings (Reumers et al., 2012) in which the authors showed that a large number of discordant variants were detected in

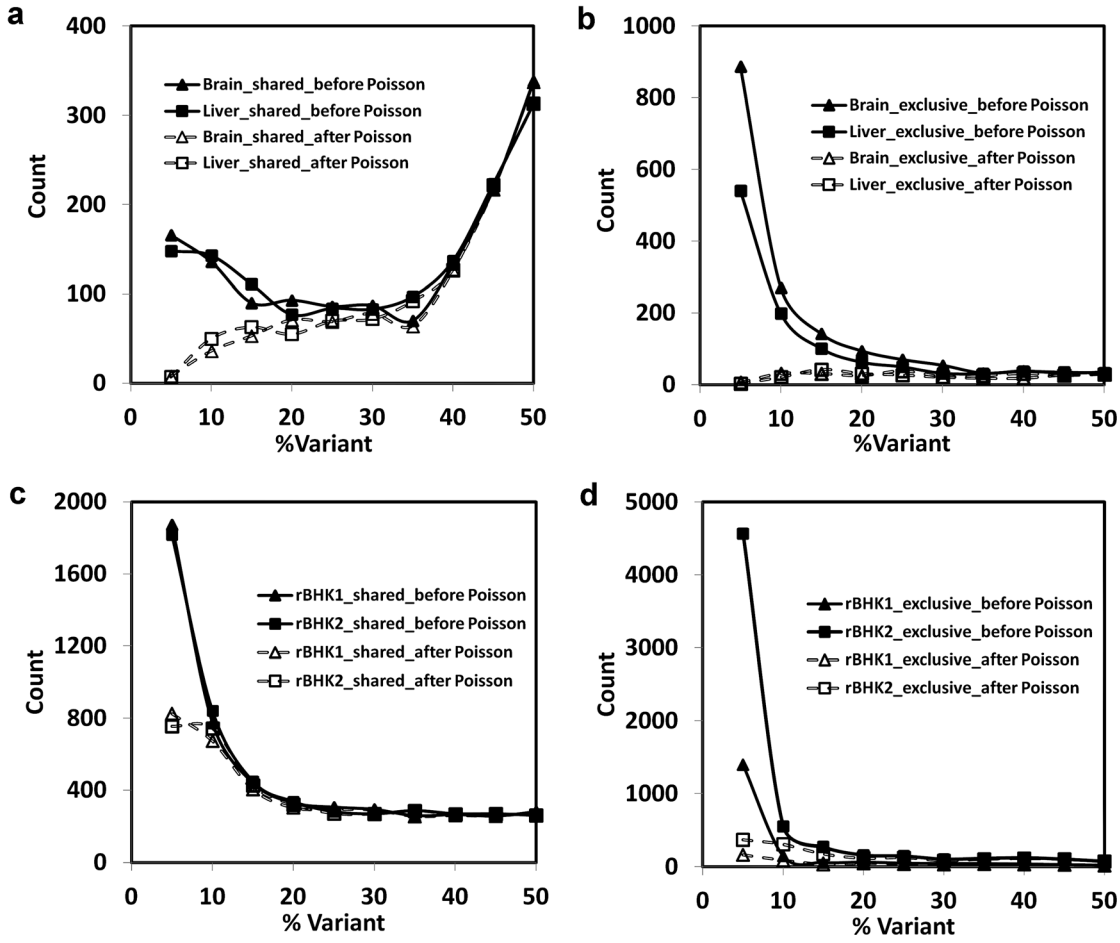

Figure 6. Distribution of variant frequencies in high RPKM contigs before and after Poisson filter in (**a**) Category 1 and 2 variants of the tissues, (**b**) Category 4 and 5 variants of the tissues, (**c**) Category 1 and 2 variants of the cell line, and (**d**) Category 4 and 5 variants of the cell line.

the same tumor DNA sample sequenced twice in two separate batches. Additionally, the greater sequencing depth in rBHK2 may contribute to the larger number of variants called in rBHK2 versus rBHK1 (Ajay et al., 2011).

Variants in Highly Expressed Genes

Genes that are highly expressed give rise to a larger number of reads, enabling detection of rare transcript variants with high confidence. We investigated the presence of variants in the transcripts of a few high-coverage contigs (Supplementary Table SV). All potential variants detected in these contigs upon application of the previously discussed filters are shown in Table IIIa and b.

From the liver and brain libraries, transcript variants were detected in several of the selected contigs. The detected variants on the two contigs CMAU029538 (Gapdh) and CMAU042624 (Rps27a) were located at the same nucleotide positions in both the liver and the brain libraries. They are likely to be heterozygous variants as suggested by their frequencies (∼50%) (Table IIIa, bold and italicized). The variants in the contig annotated to the mitochondrial genome are also located in the same positions in both tissue libraries. For rBHK libraries, in contrast, no variant was detected in any contig except CMAU221473, which has one variant site detected from rBHK2 reads at a read frequency of 2.4% (Table IIIb). The mitochondrial genome contig contains 3 and 6 potential variant sites in rBHK1 and rBHK2, respectively.

In the course of detecting the variant in CMAU221473, several candidate single nucleotide insertion/deletion transcript variants were also identified in this contig (data not shown). Sanger sequencing was performed to verify the presence of a potential G-to-A (G → A) single nucleotide substitution (read frequency of 2.4% in rBHK2) and a potential insertion of an A single nucleotide base (read frequency of 4.6% in rBHK1, 11.4% in rBHK2) in this contig. 4/90 (4.4%) *E. coli* clones sequenced reveal the presence of the single nucleotide substitution, while 7/ 88 (7.9%) clones verified the presence of the single nucleotide insertion in the cDNAs generated using the SSIII RT (Table IV, Supplementary Figs. S7 and S8). However, neither variant was found in the genomic DNA by Sanger sequencing, suggesting the possibility that the variants were introduced by the reverse transcription process (Table IV).

Table IV. Verification of potential variants in cDNA and gDNA by Sanger sequencing.

		#Variant clone/#Total clone sequenced		
Type	Nucleotide change	cDNA (SSIII RT)	cDNA (M-MuLV RT)	Genomic DNA
Substitution	G → A	4/90	0/186	0/191
Insertion	+A	7/88	1/2	0/70

Reverse transcription has several well-known limitations including template switching (Cocquet et al., 2006; Mader et al., 2001), self-priming due to mRNA secondary structure (Haddad et al., 2007), and lack of a 3′ → 5′ exonuclease proofreading mechanism (Roberts et al., 1989). We used two different reverse transcriptases in our cDNA preparation: the SSIII RT (with non-functional RNase H domain) and the M-MuLV RT (with fully functional RNase H

Table III. Variants detected among highly abundant genes.

Contig	Gene name	Position	Consensus	Variant	Variant frequency in the liver library	Variant frequency in the brain library
(a) In the liver and the brain						
CMAU029538	***Gapdh***	***191***	***A***	***G***	***47.9%***	*49.4%*
CMAU042624	***Rps27a***	***142***	***A***	***G***	***47.5%***	*47.6%*
CMAU042624	***Rps27a***	***223***	***T***	***C***	***14.6%***	*8.4%*
CMAU042624	***Rps27a***	***253***	***T***	***C***	***21.5%***	*11.1%*
CMAU042624	***Rps27a***	***265***	***A***	***G***	***5.2%***	*3.0%*
CMAU042624	Rps27a	277	G	C	3.3%	Not detected
CMAU042624	Rps27a	278	T	C	3.2%	Not detected
CMAU221474	***Mitochondria***	***4,936***	***A***	***G***	***25%***	*37.2%*
CMAU221474	***Mitochondria***	***7,143***	***G***	***A***	***30%***	*25.5%*

Contig	Gene name	Position	Consensus	Variant	Variant frequency in rBHK1	Variant frequency in rBHK2
(b) In rBHK1 and rBHK2						
CMAU221473	N/A	3,140	G	A	Not detected	2.4%
CMAU221474	Mitochondria	1,037	A	G	Not detected	8.7 %
CMAU221474	Mitochondria	3,834	G	A	Not detected	19.6%
CMAU221474	Mitochondria	4,936	G	A	Not detected	27.8%
CMAU221474	***Mitochondria***	***7,448***	***G***	***A***	***15.2%***	*16.7%*
CMAU221474	Mitochondria	9,659	T	C	2.1%	Not detected
CMAU221474	***Mitochondria***	***115,38***	***C***	***T***	***9.5%***	*14.33%*
CMAU221474	Mitochondria	15,266	G	A	Not detected	6.7%

domain). The activity of the RNase H domain of a reverse transcriptase has been shown to reduce self-priming, although it is reported to enhance template switching. Table IV shows that the single base substitution was found only in the cDNA prepared from SSIII RT (4/90 clones) and not when M-MuLV RT was used (0/186 clones). However, the single base insertion persists in the cDNA prepared from both SSIII RT (7/88 clones) and M-MuLV RT (1/2 clones), possibly due to the lack of the $3' \rightarrow 5'$ exonuclease proofreading activity in both reverse transcriptases. The fact that false variants introduced by a faulty reverse transcriptase were detected in deep sequencing reads highlights the needs for further improvements in the current RNA-seq technology.

Mutations in Growth Regulatory Pathways

During the process of cell line derivation from a tissue, a number of mutations may arise including single nucleotide mutations. Some of these mutations may confer the cells the ability to grow without senescence in culture. Mutations in the genes of the MAPK, Insulin, PI3K/AKT, p53, and MTOR pathways may lead to dysregulation of growth control and have been linked to various cancers (Brosh and Rotter, 2009; Dhillon et al., 2007; Engelman, 2009; Mello and Attardi, 2013; Shaw and Cantley, 2006; Yuan and Cantley, 2008). We searched the genes of the MAPK, Insulin, PI3K/AKT, MTOR growth signaling pathways for variants that are present in both the cell line data but are absent in both tissues or vice versa to probe for mutations that may confer the capability to grow in culture. We discovered 75 potentially mutated sites distributed across 27 distinct genes. All the identified mutations are either heterozygous (∼50% frequency) or homozygous (∼100% frequency). 35 of which are located in the protein coding regions and seven cause amino acid changes in the genes Mdm2, Igf2r, Map3k6, Map3k14, Pik3cb, and Araf (Supplementary Table SVI). In humans, one missense mutation in Mdm2 has been identified that reduces the activity of p53 and accelerates tumor formation (Bond et al., 2004). Several Igf2r missense mutations have been found in various human cancers (Byrd et al., 1999; Kong et al., 2000). A single somatic mutation in Map3k6 occurs recurrently in gastric cancer (Zang et al., 2011). Genetic aberrations in Map3k14 including single nucleotide mutations activate the NF-$_K$B pathway in patients with Multiple Myeloma (MM) (Annunziata et al., 2007; Keats et al., 2007; Rossi et al., 2011). Mutations in the Pik3cb and Araf fail to elicit oncogenic transformation (Ciraolo et al., 2008; Jia et al., 2009; Mercer et al., 2002). It is conceivable that some of these detected mutations may have played a role in derivation of the BHK cell line from the tissue.

Discussion

We have employed high-throughput sequencing and the Oases transcriptome assembler to create a BHK cell line transcriptome augmented by Syrian hamster tissue data. Previous experience with CHO transcriptome assembly efforts indicates that inclusion of sequence reads from diverse tissues or culture conditions can significantly augment an assembly by providing higher coverage of transcripts expressed at low levels in a single cell line or tissue (Jacob et al., 2010). For example, Pklr, the liver and red blood cell isozyme of pyruvate kinase, is expressed at 110 RPKM in liver but only 0.01 RPKM in rBHK1 (Fig. 2) and 0.2 RPKM in brain (not shown); it is thus likely that the liver library provided the majority of reads used to assemble this gene. In addition, several thousand contigs were built solely from the tissue reads, many of which represent genes with known tissue-specific expression. Given that (1) about 80% of reads in the rBHK libraries map to the assembly, (2) the assembly already represents >15,000 unique genes, and (3) >90% of contigs are mapped by rBHK reads, the majority of genes expressed in this cell line are probably already represented in the assembly. Adding a new cell type or condition would likely be needed to substantially improve the number of genes assembled.

This assembly provides a powerful tool for BHK or Syrian hamster transcriptome analysis. Primers, siRNAs, and constructs for gene knockouts can be designed without reliance on nucleotide identity to other species. Using RNA-seq or microarray studies to generate whole transcriptome profiles will facilitate the formulation of new hypotheses. For example, in rBHK we saw evidence that the dominant glycolytic isozymes were those typical of proliferating cells, while the isozyme levels in the tissues were consistent with a quiescent phenotype.

We investigated the ability of high-throughput RNA sequencing technology to detect the presence of rare transcript variants. Using stringent filtering criteria, we detected high-confidence variant candidates in several abundant transcripts in rBHK. We verified the existence of one single nucleotide substitution and one single nucleotide insertion in the cDNA of a gene contig using Sanger technology. However, the same variants were not present in the Sanger-sequenced genomic DNA. It is possible that there are multiple copies of this gene in the rBHK cells, with the variants being expressed at a moderate level from one copy. As a result, we could detect the variants only at a transcript level, but not at a genomic level.

Detection of sequence variants at the transcriptome level may complement genomic studies, for example, by showing whether a mutation which confers selective advantage is actually expressed in the transcripts. Here, we showed several genes in the growth signaling pathways are potentially mutated in the cell line. Indeed some of these mutations may cause activation of the cell cycle genes as shown by the transcript levels.

We have created a reference transcriptome for a BHK cell line, thereby removing an obstacle to many other possible studies. When the newly released Syrian hamster genome is annotated, providing complementary information such as gene structure and organization, even more doors will be opened to researchers to develop new tools and knowledge of

BHK and Syrian hamster transcriptomic and genomic characteristics.

The authors thank Dr. Robert Meisel of the Department of Neuroscience at the University of Minnesota for his kind gift of the Syrian hamster tissues. Thanks also to Terk Shuen Lee and Faraaz Yusufi for advice on read pre-processing and contig annotation. Computational resources and support were provided by the Minnesota Supercomputing Institute. The study was funded by Bayer HealthCare. K.C.J. was supported in part by an NSF Graduate Research Fellowship.

References

Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. 2010. A map of human genome variation from population-scale sequencing. Nature 467:1061–1073.

Ajay SS, Parker SC, Abaan HO, Fajardo KV, Margulies EH. 2011. Accurate and comprehensive sequencing of personal genomes. Genome Res 21:1498–1505.

Annunziata CM, Davis RE, Demchenko Y, Bellamy W, Gabrea A, Zhan F, Lenz G, Hanamura I, Wright G, Xiao W, et al. 2007. Please provide complete list of author names for all the et al. references.Frequent engagement of the classical and alternative NF-kappaB pathways by diverse genetic abnormalities in multiple myeloma. Cancer cell 12: 115–130.

Bhathena J, Kulamarva A, Martoni C, Urbanska A, Malhotra M, Paul A, Prakash S. 2011. Diet-induced metabolic hamster model of nonalcoholic fatty liver disease. Diabetes Metab Syndr Obes Targets Ther 4:195–203.

Bond GL, Hu W, Bond EE, Robins H, Lutzker SG, Arva NC, Bargonetti J, Bartel F, Taubert H, Wuerl P, et al. 2004. A single nucleotide polymorphism in the MDM2 promoter attenuates the p53 tumor suppressor pathway and accelerates tumor formation in humans. Cell 119:591–602.

Brosh R, Rotter V. 2009. When mutants gain new powers: News from the mutant p53 field. Nat Rev Cancer 9:701–713.

Bullard JH, Purdom E, Hansen KD, Dudoit S. 2010. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinform 11:94.

Butler M. 2006. Optimisation of the cellular metabolism of glycosylation for recombinant proteins produced by mammalian cell systems. Cytotechnology 50:57–76.

Byrd JC, Devi GR, de Souza AT, Jirtle RL, MacDonald RG. 1999. Disruption of ligand binding to the insulin-like growth factor II/mannose 6-phosphate receptor by cancer-associated missense mutations. J Biol Chem 274:24408–24416.

Ciraolo E, Iezzi M, Marone R, Marengo S, Curcio C, Costa C, Azzolino O, Gonella C, Rubinetto C, Wu H, et al. 2008. Phosphoinositide 3-kinase p110β activity: Key role in metabolism and mammary gland cancer but not development. Sci Signal 1:1–12.

Cocquet J, Chong A, Zhang G, Veitia RA. 2006. Reverse transcriptase template switching and false alternative transcripts. Genomics 88:127–131.

Crespo MJ, Escobales N. 2008. Early pathophysiological alterations in experimental cardiomyopathy: The Syrian cardiomyopathic hamster. P R Health Sci J 27:307+.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis TAA, del Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491–498.

Dhillon AS, Hagan S, Rath O, Kolch W. 2007. MAP kinase signalling pathways in cancer. Oncogene 26:3279–3290.

Dillard A, Matthan N, Lichtenstein A. 2010. Use of hamster as a model to study diet-induced atherosclerosis. Nutr Metab 7:89.

Dingermann T. 2008. Recombinant therapeutic proteins: Production platforms and challenges. Biotechnol J 3:90–97.

Engelman JA. 2009. Targeting PI3K signalling in cancer: Opportunities, challenges and limitations. Nat Rev Cancer 9:550–562.

Glenn TC. 2011. Field guide to next-generation DNA sequencers. Mol Ecol Resour 11:759–769.

Haddad F, Qin AX, Giger JM, Guo H, Baldwin KM. 2007. Potential pitfalls in the accuracy of analysis of natural sense-antisense RNA pairs by reverse transcription-PCR. BMC Biotechnol 7:21.

Jacob NM, Kantardjieff A, Yusufi FNK, Retzel EF, Mulukutla BC, Chuah SH, Yap M, Hu W-S. 2010. Reaching the depth of the Chinese hamster ovary cell transcriptome. Biotechnol Bioeng 105:1002–1009.

Jia S, Roberts TM, Zhao JJ. 2009. Should individual PI3 kinase isoforms be targeted in cancer? Curr Opin Cell Biol 21:199–208.

Jiang R, Monroe T, McRogers R, Larson PJ. 2002. Manufacturing challenges in the commercial production of recombinant coagulation factor VIII. Haemophilia 8:1–5.

Jové M, Ayala V, Ramírez-Núñez O, Serrano JCE, Cassanyé A, Arola L, Caimari A, del Bas JM, Crescenti A, Pamplona R, et al. 2012. Lipidomic and metabolomic analyses reveal potential plasma biomarkers of early atheromatous plaque formation in hamsters. Cardiovasc Res.Please provide volume number and page range.

Keats JJ, Fonseca R, Chesi M, Schop R, Baker A, Chng WJ, Van Wier S, Tiedemann R, Shi CX, Sebag M, et al. 2007. Promiscuous mutations activate the noncanonical NF-kappaB pathway in multiple myeloma. Cancer Cell 12:131–144.

Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 22:568–576.

Kong F-M, Anscher MS, Washington MK, Killian JK, Jirtle RL. 2000. M6P/IGF2R is mutated in squamous cell carcinoma of the lung. Oncogene 19:1572–1578.

Lam HY, Clark MJ, Chen R, Natsoulis G, O'Huallachain M, Dewey FE, Habegger L, Ashley EA, Gerstein MB, Butte AJ, et al. 2012. Performance comparison of whole-genome sequencing platforms. Nat Biotechnol 30:78–82.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrowsâ"Wheeler transform. Bioinformatics 25:1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project, Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

Macpherson I, Stoker M. 1962. Polyoma transformation of hamster cell clones—An investigation of genetic factors affecting cell competence. Virology 16:147–151.

Mader RM, Schmidt WM, Sedivy R, Rizovski B, Braun J, Kalipciyan M, Exner M, Steger GG, Mueller MW. 2001. Reverse transcriptase template switching during reverse transcriptase-polymerase chain reaction: Artificial generation of deletions in ribonucleotide reductase mRNA. J Lab Clin Med 137:422–428.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303.

Mello SS, Attardi LD. 2013. Not all p53 gain-of-function mutants are created equal. Cell Death Differ 20:855–857.

Mercer K, Chiloeches A, Huser M, Kiernan M, Marais R, Pritchard C. 2002. ERK signalling and oncogene transformation are not impaired in cells lacking A-Raf. Oncogene 21:347–355.

Meyer N, Penn LZ. 2008. Reflecting on 25 years with MYC. Nat Rev Cancer 8:976–990.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Meth 5:621–628.

Paessler S, Aguilar P, Anishchenko M, Wang H-Q, Aronson J, Campbell G, Cararra A-S, Weaver SC. 2004. The hamster as an animal model for eastern equine encephalitis—And its use in studies of virus entrance into the brain. J Infect Dis 189:2072–2076.

Pay TW, Boge A, Menard FJ, Radlett PJ. 1985. Production of rabies vaccine by an industrial scale BHK 21 suspension culture process. Dev Biol Stand 60:171–174.
Popov D, Simionescu M, Shepherd PR. 2003. Saturated-fat diet induces moderate diabetes and severe glomerulosclerosis in hamsters. Diabetologia 46:1408–1418.
Radlett PJ, Pay TW, Garland AJ. 1985. The use of BHK suspension cells for the commercialization production of foot and mouth disease vaccines over a twenty year period. Dev Biol Stand 60:163–170.
Requena JM, Soto M, Doria MD, Alonso C. 2000. Immune and clinical parameters associated with Leishmania infantum infection in the golden hamster model. Vet Immunol Immunopathol 76:269–281.
Reumers J, De Rijk P, Zhao H, Liekens A, Smeets D, Cleary J, Van Loo P, Van Den Bossche M, Catthoor K, Sabbe B, et al. 2012. Optimized filtering reduces the error rate in detecting genomic variants by short-read sequencing. Nat Biotechnol 30:61–68.
Roberts JD, Preston BD, Johnston LA, Soni A, Loeb LA, Kunkel TA. 1989. Fidelity of two retroviral reverse transcriptases during DNA-dependent DNA synthesis in vitro. Mol Cell Biol 9:469–476.
Rossi D, Deaglio S, Dominguez-Sola D, Rasi S, Vaisitti T, Agostinelli C, Spina V, Bruscaggin A, Monti S, Cerri M, et al. 2011. Alteration of BIRC3 and multiple other NF-kappaB pathway genes in splenic marginal zone lymphoma. Blood 118:4930–4934.
Shaw RJ, Cantley LC. 2006. Ras, PI(3)K and mTOR signalling controls tumour cell growth. Nature 441:424–430.
Staffend NA, Meisel RL. 2012. Aggressive experience increases dendritic spine density within the nucleus accumbens core in female syrian hamsters. Neuroscience 227:163–169.
Stoker M, Macpherson IAN. 1964. Syrian hamster fibroblast cell line BHK21 and its derivatives. Nature 203:1355–1357.
Xiao S-Y, Guzman H, Zhang H, Travassos da Rosa APA, Tesh RB. 2001. West Nile virus infection in the golden hamster (*Mesocricetus auratus*): A model for west nile encephalitis. Emerg Infect Dis 7:714.
Yee JC. 2008. Genomic and proteomic profiling of mammalian cells under high productivity states. in chemical engineering and materials science. Minneapolis, MN: University of Minnesota-Twin Cities.
Yuan TL, Cantley LC. 2008. PI3K pathway alterations in cancer: Variations on a theme. Oncogene 27:5497–5510.
Zang ZJ, Ong CK, Cutcutache I, Yu W, Zhang SL, Huang D, Ler LD, Dykema K, Gan A, Tao J, et al. 2011. Genetic and structural variation in the gastric cancer kinome revealed through targeted deep sequencing. Cancer Res 71:29–39.

Supporting Information

Additional supporting information may be found in the online version of this article.