A faster workflow for the assessment of genomic loci in mice using a novel HMW DNA extraction technology upstream of Cas9 targeted sequencing

Simon Lesbirel[1], Ph.D., Jeremy R. Charette[1], B.S., Chia-Lin Wei[1], Ph.D., Eric Cantor[2], Ph.D., Danielle Freedman[2], M.S., Giron Koetsier[2], Ph.D.

1: The Jackson Laboratory, Bar Harbor, ME/Farmington, CT and 2: New England Biolabs, Inc., Ipswich, MA

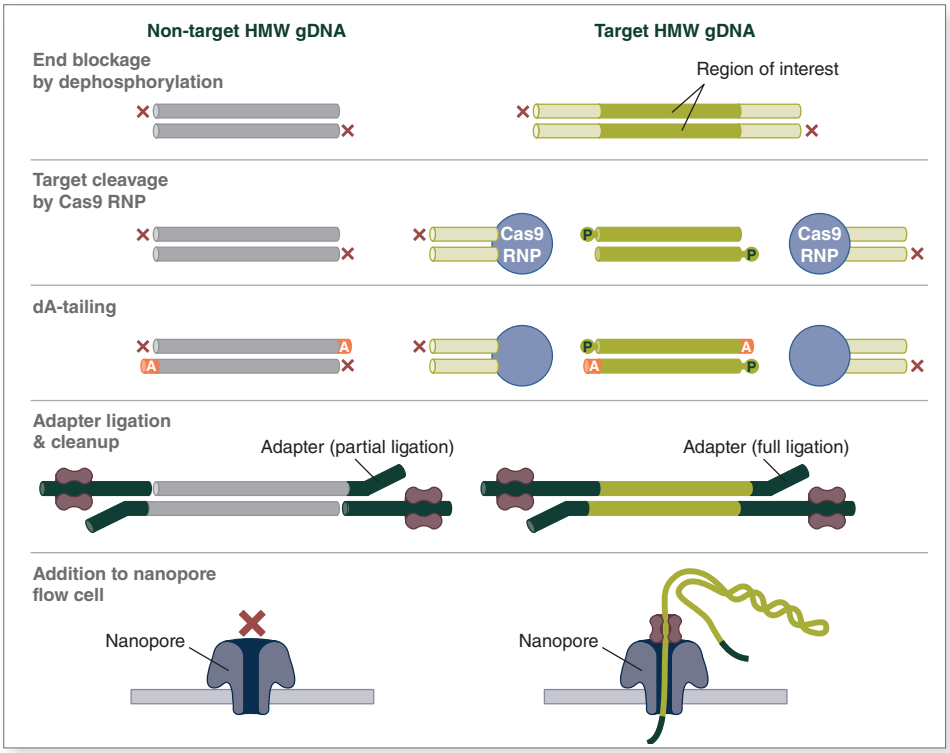
INTRODUCTION

Generation of transgenic mice through random or targeted integration of DNA fragments can lead to structural variation and integration mutagenesis (1), both of which are undesirable outcomes. Due to the significant labor required for their characterization, it is estimated that only around 5% of transgenic mouse models published in the Mouse Genome Database have an annotated chromosomal location (1). Therefore, a technique capable of quickly and cost-effectively identifying chromosomal location and confirming the transgene sequence integrity is essential. Further to this, the interrogation of large loci at the base level between strains remains difficult without using whole genome sequencing. A recently described technique, Cas9 no-amplification enrichment (2), has the potential to fulfill that need.

Traditionally, the genomic modifications required to generate mouse models leverage PCR-based assays and Sanger sequencing for validation. However, in many cases, the structure and the sequence of the gene or its chromosomal integration site hinder analysis by these methods. Loops and sequence repeats prevent effective assessment of DNA sequence. The CRISPR/Cas9-mediated amplification-free enrichment approach for Oxford Nanopore Technologies® sequencing is an alternative method for interrogation of loci of interest or transgene sites. The method is relatively low-cost and can enrich regions of interest over native sequences without the need for PCR amplification (Figure 1).

In standard ligation-based whole genome sequencing approaches, desired loci/transgenes will be sequenced only once or a few times per nanopore sequencing run, but not with enough coverage to collect reliable sequence information. The Cas9 no-amplification enrichment workflow allows for specific enrichment of targeted regions by reducing undesired fragments from the sequencing process via dephosphorylation of their phosphate ends. Lacking terminal 5´ phosphate groups, they do not participate in adapter ligation. The target region, however, is subsequently cleaved using a Cas9-sgRNA (single guide RNA) ribonucleoprotein complex (RNP) making it accessible for sequencing adapter ligation. The resulting libraries allow for enriched sequence generation from the region of interest against a minimal background of genomic DNA sequences typically resulting from off-target Cas9 cleavage and non-specific adapter ligation. Furthermore, multiple sgRNAs can be used to enrich a variety of targets in a single library, thereby increasing efficiency and decreasing cost (3).

FIGURE 1: **Overview of the Cas9 no-amplification enrichment library prep workflow**



At The Jackson Laboratory, one of our priorities has been to establish assays to standardize analysis for routine assessment of genomic alterations such as targeted mutagenesis and transgene integrations. Cas9 enrichment has proven to be an effective approach. Simultaneous Cas9 enrichment analysis of 2 to 4 targeted sequences has now been established as a standard workflow, with targeted regions typically being around 5 kb in size and sometimes up to 30 kb.

Successful use of the Cas9 enrichment protocol relies on using high-quality, high molecular weight (HMW) gDNA as an input material. Working with the longest possible DNA fragments increases the chance that the entire region of interest remains intact after DNA extraction. The initial Cas9 enrichment sequencing workflow implemented within the Genome Technologies group at The Jackson Laboratory was dependent on phenol/chloroform DNA extraction, which initially fulfilled the requirements. However, while the phenol/chloroform-based workflow is effective for ultralong sequencing, it proved to be laborious and time consuming when applied to the Cas9 enrichment protocol. The sample lysis, phenol extraction and DNA precipitation take approximately one full day. Subsequently, this method requires up to 3 days of "rest time" to allow the isolated HMW gDNA to return to solution, resulting in the whole extraction process taking several days. Accordingly, the Cas9 enrichment can be started around day 5 (Figure 8). In addition, the increased frequency of extractions produced excessive amounts of hazardous waste. Therefore, we sought a faster and more environmentally friendly DNA extraction alternative.

In this work, we leveraged the novel glass bead-based approach employed by the New England Biolabs Monarch® HMW DNA Extraction Kits to significantly reduce the time required for generating HMW gDNA from mouse tissue samples. With this new approach, the HMW DNA extraction process from tissues is complete in about 90 minutes, and DNA is ready to use shortly after, thereby significantly reducing the overall time required to perform the Cas9 enrichment workflow. Yield, purity, and integrity of the isolated HMW DNA is compared to phenol-extracted DNA and its efficient use in the optimized Cas9 sequencing workflow is demonstrated. Sequencing results are presented describing several case studies: locus analysis for mouse strain comparison, targeted mutagenesis sites and several transgenes and their insertion sites, including loci that proved difficult to sequence in earlier attempts.

METHODS

Description of samples and inserts

Transgenic mice with homozygous and heterozygous inserts as well as targeted mutations and endogenous regions of the genome (5 kb to 30 kb) were analyzed (Table 1). Liver and kidney serve as good sources of HMW DNA and were therefore chosen as target organs. Organs were harvested in accordance with the ethical standards at The Jackson Laboratory, and samples were flash frozen in liquid nitrogen and pulverized prior to extraction.

HMW DNA extraction with phenol/chloroform

HMW gDNA was extracted using a modified phenol/chloroform extraction protocol (4). 10–20 mg of tissue was pulverized on dry ice, incubated in 10 ml lysis buffer (100 mM NaCl; 10 mM Tris, pH 8.0; 25 mM EDTA, pH 8.0; 0.5% SDS; 0.02 mg/ml RNase A) at 37°C for 1 hour. After addition of Proteinase K (0.1 mg/ml final concentration), samples were incubated at 50°C for 3 hours to complete tissue digestion. DNA was extracted using phenol:chloroform:isoamyl alcohol (25:24:1, v/v), phase separated twice using Qiagen® MaXtract™ High Density tubes, centrifuging at 3000 x *g* for 10 minutes, precipitated in ice cold ethanol, washed twice in 70% ethanol, and resuspended in 400 µl TE low. The purified DNA rested at 4°C for 2–3 days to ensure complete dissolving.

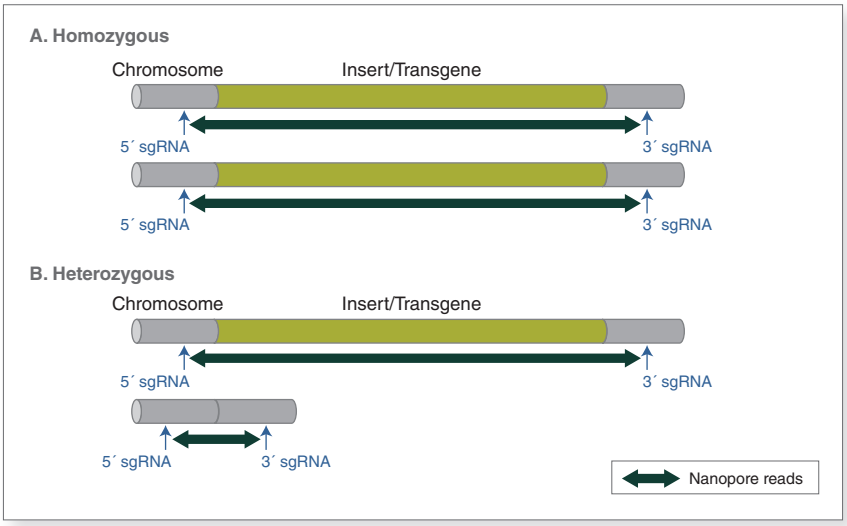
HMW DNA extraction using Monarch HMW DNA Extraction Kits

HMW genomic DNA from mouse tissue samples was isolated using the Monarch HMW DNA Extraction Kit for Tissue (NEB #T3060) using the recommended protocols. For extraction from mouse liver samples, the recommended addition of NaCl to a final concentration of 2.2M before bead binding was included in the tissue protocol. Agitation speed used during lysis was either 500 rpm and 1700–2000 rpm to generate longer and shorter fragments, respectively. Following extraction, if DNA was not homogenously dissolved, DNA was prepared for use by heating to 50°C for 10 minutes with occasional pipetting up and down with a P200 wide bore tip. If extractions were done at the end of the day, DNA samples were stored at 4°C overnight prior to analysis and library prep on the following day.

TABLE 1: **Description of target regions investigated**

MOUSE LINE NUMBER	REGION OF INTEREST SIZE (kb)	ORGAN TYPE
1	13	Mouse liver
4	5	
5	30	
6	5	Mouse kidney
7	5	
8	5	
9	5	
10	19	
11	22.5	

FIGURE 2: **Depiction of sgRNA targeting design for homozygous and heterozygous transgene insertions**



DNA measurement and analysis

A rapid screen of DNA integrity (DIN) was carried out by TapeStation® 4200 (Agilent Technologies) using Genomic DNA ScreenTapes® and reagents (Agilent Technologies #5067-5365). Rapid assessment of the amount of HMW DNA (>50 kb) was carried out by region analysis of the DNA signals on the TapeStation. DNA Purity was assessed by Nanodrop® 2000 and yield was analyzed with both Nanodrop and Qubit® dsDNA BR Assay Kit and the Qubit Fluorometer 3.0 (Thermo Scientific). For concentration measurement, 1 µl of DNA was taken from the top, middle, and bottom of the HMW DNA sample and averaged.

Cas9 enrichment library generation

The Cas9 sequencing protocol was carried out using Oxford Nanopore's Ligation Sequencing Kit (SQK-LSK109) coupled with the New England Biolabs NEBNext® Companion Module (NEB #E7180). Alternatively, this protocol can be conducted using a specific Cas9 Sequencing Kit (Oxford Nanopore Technologies SQK-CS9109). The corresponding Oxford Nanopore Technologies protocol was followed for library construction with minor modifications as described below. sgRNAs were designed to avoid common SNP sites (5) and were sourced from Integrated DNA Technologies (IDT), along with the described Cas9 enzyme and duplex buffer. Typical sgRNA design for homozygous and heterozygous inserts is shown in Figure 2. A few protocol modifications were implemented to improve the relatively low efficiency of the Cas9 library preparation caused by the high viscosity of the HMW samples:

- Cas9-cleavage was carried out for 1 hour
- Sequencing adaptor ligation was carried out for 1 hour instead of 10 minutes
- Elution incubation time was 30 minutes instead of 10 minutes, and elution was carried out in 13 µl (1 µl for measurement, 12 µl for sequencing)

Oxford Nanopore Technologies sequencing

Samples were sequenced on MinION R9.4.1 flow cells for 24 hours on either MinION MK1B or GridION Mk1 (Oxford Nanopore Technologies). Samples were run as single runs or multiplexed with up to 4 targets. Flow cells were reused 2 to 4 times after washing every 24 hours using the protocol and reagents from Oxford Nanopore Technologies (Flow Cell Wash Kit, EXP-WSH003).

Sequencing data analysis

Base calling was carried out using GUPPY (v3.2.10). The resulting fastq files were aligned to a reference sequence using minimap2 (v2.17). Custom reference sequences were constructed for transgene insertions sites using the *Mus musculus* reference genome (MM10) and corresponding insert sequence. Alignment results were subject to MapQ score filtering using Samtools (v1.11). Subsequent coverage depth for on-target reads was generated using Bedtools (v2.29.2). On-target reads were visualized in Integrative Genomics Viewer (IGV).

RESULTS

Assessment of DNA concentration, purity and fragment length

The HMW DNA that was used for initial Cas9 enrichment studies was isolated using phenol/chloroform extraction. For later studies, the Monarch HMW DNA Extraction Kit for Tissue was introduced. Yield and purity data are shown for DNA extracted from liver and kidney samples with phenol/chloroform (Table 2) or Monarch (Table 3).

Phenol/chloroform-extracted samples

Concentration and purity measurements by Nanodrop and Qubit (Table 2) revealed that based on spectrophotometric readings, around 1.6 µg DNA per mg tissue was isolated for liver and 4-6 µg per mg tissue for kidney. However, Qubit values were significantly lower, particularly for the liver samples. Purity grade was intermediate; though A_{260}/A_{280} purity ratios were in the normal range (1.83-1.85), the A_{260}/A_{230} ratios were lower than optimal, with 1.8-2.1 for kidney and only 1.1-1.3 for liver.

Monarch HMW DNA Extraction Kit-extracted samples

Concentration and purity measurements were assessed by Nanodrop and Qubit (Table 3). Yields of initial samples were somewhat lower than expected, but good yields (1.4 µg/mg for liver and 2–3 µg/mg for kidney) were obtained after becoming familiar with the protocol. Consistent high purity was observed among all extracted samples as A_{260}/A_{280} purity ratios were greater than 1.83 and A_{260}/A_{230} ratios were >2.1.

Rapid assessment of DNA integrity

For a quick assessment of the fragment length of the phenol/chloroform- and the Monarch- extracted DNA samples, samples were run on the TapeStation using Genomic DNA ScreenTapes. With the help of the region analysis tool the percentage of HMW DNA (DNA > 50 kb) was determined. Overall, the percentages were 69% for the phenol/chloroform samples and around 81% for the Monarch samples (Table 4). A typical example of such region analysis is shown in Figure 3. DIN values were comparable for both extraction methods.

Cas9-targeted sequencing metrics and coverage

Cas9-targeted sequencing was carried out on a variety of genomic loci, including transgenic insertions and sites of targeted mutagenesis. Sequencing coverage metrics are shown in Table 5. Although the percentage of on-target reads is low compared to total reads generated, the mean coverage of targeted regions is far greater than the whole genome coverage generated from each run, as shown in Table 6.

TABLE 2: **Yield and purity analysis of phenol/chloroform-extracted HMW DNA samples**

HMW DNA samples were analyzed on a Nanodrop 2000 (Thermo Scientific) to determine the concentration and purity ratios. Additionally, fluorometric concentration assessment was carried out using a Qubit dsDNA BR Assay Kit and a Qubit 3.0 fluorometer (both Thermo Scientific). In each case 1 µl samples were taken from top, middle and bottom of the solution. Averages of the three were calculated and displayed in the table.

SAMPLE INFORMATION		EXTRACTION PROTOCOL DETAILS		CONCENTRATION		YIELD		PURITY	
SAMPLE	TISSUE TYPE	INPUT AMOUNT (mg)	ELUATE VOLUME (µl)	QUBIT CONCENTRATION (ng/µl)	NANODROP CONCENTRATION (ng/µl)	YIELD (µg)	YIELD PER MG TISSUE (µg)	A260/A280	A260/A230
A	Liver	15	400	72.6	241.2	29.0	1.9	1.85	1.33
B	Liver	15	400	53.5	230.7	21.4	1.4	1.84	1.12
C	Kidney	20	400	294.0	398.7	117.6	5.9	1.83	2.08
D	Kidney	20	400	210.0	268.2	84.0	4.2	1.86	1.77
E	Kidney	20	400	188.0	280.9	75.2	3.8	1.85	2.07

TABLE 3: **Yield and purity analysis of HMW DNA samples extracted with the Monarch HMW DNA Extraction Kit**

HMW DNA samples were analyzed on a Nanodrop 2000 to determine the concentration and the purity ratios. Additionally, fluorometric concentration assessment was carried out using a Qubit dsDNA BR Assay Kit and a Qubit 3.0 fluorometer. In each case 1 µl samples were taken from top, middle and bottom of the solution. Averages of the three were calculated and are displayed in the table. DNA samples from kidney purified using 500 rpm agitation speed during lysis showed some loss of yield; the reduced gDNA yield from 500 rpm agitated samples can be mitigated by increasing bead binding incubation from 5 minutes (yellow) to 8 minutes (green).

SAMPLE INFORMATION		EXTRACTION PROTOCOL DETAILS			CONCENTRATION		PURITY		YIELD	
MOUSE LINE	SAMPLE TYPE	INPUT (mg)	LYSIS AGITATION SPEED (RPM)	BINDING TIME ON ROTATOR (MIN)	QUBIT CONCENTRATION (ng/µl)	NANODROP CONCENTRATION (ng/µl)	A260/A280	A260/A230	TOTAL YIELD (µg)	YIELD PER MG (µg)
1	Liver	24	2000	5	420.5	479.3	1.81	2.23	21	0.9
2	Liver	12	2000	5	54.1	60	1.83	2.21	5.4	0.5
3	Liver	16	2000	5	48.4	56.3	1.84	2.26	4.8	0.3
4	Liver	15	1700	5	212	229	1.86	2.27	21.2	1.4
5	Liver	10	500	5	132	174.7	1.84	2.14	13.2	1.3
6	Kidney	17	2000	5	338	337	1.85	2.27	33.8	2
7	Kidney	20	2000	5	547	592.5	1.83	2.25	54.7	2.7
8	Kidney	10	500	5	52.8	49	1.87	2.09	5.3	0.5
9	Kidney	9	500	5	56.7	66.9	1.86	2.2	5.7	0.6
10	Kidney	12	500	8	163.5	195.3	1.85	2.35	16.4	1.4
11	Kidney	10	500	8	193	197.9	1.84	2.36	19.3	1.9
12	Kidney	10	500	8	211	227	1.85	2.34	21.1	2.1
13	Kidney	7	500	5	32.7	34.3	1.89	2.25	3.3	0.5
14	Kidney	14	500	5	121.3	165	1.87	2.24	12.1	0.9

TABLE 4: **HMW DNA samples isolated using the Monarch Kit demonstrate longer fragment legnths than phenol/ chloroform extraction**

1 µl HMW DNA samples were loaded on Genomic DNA ScreenTapes using accompanying reagents and run on a TapeStation 4200 device (Agilent Technologies) to produce DNA Integrity (DIN) values. At the end of each run a % value of DNA >50 kb was determined by using the region analysis tools and integrating the signal area above 50 kb.

EXTRACTION APPROACH	SAMPLE	DIN	DNA > 50 KB (%)	DNA > 50 KB AVG. (%)
Phenol	A	9.6	73.5	69.4
	B	9.4	71.4	
	C	9.3	81.1	
	D	9.8	75.3	
	E	9.9	45.9	
Monarch	1	9.8	77	81.2
	2	9.8	88.9	
	3	9.3	81.8	
	4	9.6	78.5	
	5	9.5	79.4	
	6	9.4	81.8	

Run metrics from the multiple 24-hour targeted sequencing runs are summarized in Table 7. Samples 1, 10 and 11 were run on Oxford Nanopore Technologies GridION platform, while all others were run on a MinION. Due to the low number of target regions, only a small proportion of DNA ligates with the sequencing adaptors, resulting in low sequencing data overall. Accordingly, only 10-15% pore usage is observed (Figure 4).

Targeted sequencing case study results

Targeted sequencing for rapid strain comparison

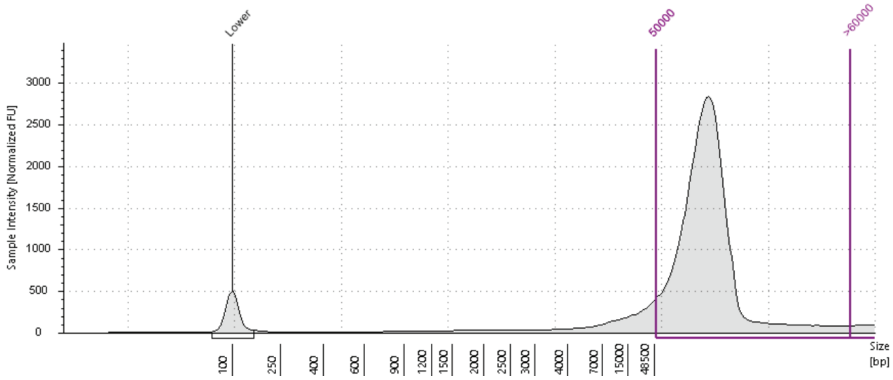
CRISPR/Cas9 enrichment sequencing was employed to investigate sequence variation at the MX1 locus in its entirety by targeting 2 kb up- and downstream. This assay was applied to the common laboratory strain C57BL/6J (Sample 10) and a wild-derived strain CAST/EiJ (Sample 11). The resulting data confirm a known 3.5kb deletion in C57BL/6J spanning exons 8 to 12. The deletion appears as an insertion in our CAST/EiJ (Sample 11) alignment due to the mouse reference being constructed with C57BL/6J. The resulting capture sequencing generated 80X coverage over a 22.5 kb region in CAST/EiJ and 230X coverage over a 19 kb region in C57BL/6J (Figure 5).

Analysis of targeted mutagenesis sites

Cas9 targeted sequencing was also employed to validate the integrity of targeted mutations within multiple mouse strains (Samples 1 and 8). In Sample 1, which harbors a 5 kb insert, 2 insertions were detected, indicated in purple in Figure 6. In Sample 8, significantly longer than expected reads were obtained (up to 95 kb) from the targeted region of 13 kb.

FIGURE 3: **Example of the HMW DNA % >50 kb region analysis of a Genomic DNA ScreenTape run**

1 µl HMW DNA samples was loaded on Genomic DNA ScreenTapes using accompanying reagents and run on a TapeStation 4200 device (all Agilent Technologies). At the end of each run a value for % of DNA >50 kb was determined by using the region analysis tools and integrating the signal area above 50 kb. An example of Monarch sample 6 is shown here.



Region Table

From [bp]	To [bp]	Average Size [bp]	Conc. [ng/µl]	Region Molarity [nmol/l]	% of Total	Region Comment	Color
50000	600000	225073	211	2.40	100.00		
50000	>60000	–	154	2.22	81.83		

TABLE 5: **Cas9-targeted sequencing coverage metrics**

Alignment and coverage data were generated for each mouse line using nanopore targeted sequencing metrics, sequencing data, and dedicated software applications as described in the methods section.

MOUSE LINE	NUMBER OF READS	NUMBER OF READS ALIGNED	NUMBER OF ON-TARGET READS	ON-TARGET PERCENTAGE	MEAN COVERAGE
1	67,863	62,794	151	0.24%	100
4	49,790	44,304	483	1.09%	92
5	49,790	44,304	74	0.17%	44
6	14,306	13,968	75	0.54%	20
7	17,017	16,648	62	0.37%	23
8	57,083	56,697	79	0.14%	70
9	57,083	56,697	211	0.37%	74
10	111,898	111,642	298	0.27%	230
11	64,636	64,303	124	0.19%	80

Analysis of transgene insert regions

Figure 7 demonstrates reads collected from mouse lines 4 and 5 with sequencing results including a homozygous 30 kb insert and a heterozygous 5 kb insert. The typical reads obtained with homozygous and heterozygous transgene inserts are illustrated in Figure 2. Similar coverage plots were obtained with Sample 9, a 5 kb insertion with 74X coverage (data not shown). Generating reads from Samples 6 and 7 (heterozygous 5 kb insert) was not successful with ultra HMW DNA. Therefore, the Monarch HMW DNA extraction was repeated with larger input amounts and lysis was carried out at maximal agitation speed (2000 rpm) to reduce the size of the HMW DNA fragments and the accompanying viscosity. This modification yielded DNA that resulted in a usable 20X/23X coverage of the regions of interest (data not shown).

TABLE 6: **Comparison of on-target coverage versus whole genome coverage**

On-target coverage was compared to the whole genome coverage for mouse line 10 and 11 to demonstrate the on-target coverage enrichment.

SAMPLE	AVERAGE ON-TARGET COVERAGE	AVERAGE WHOLE GENOME COVERAGE	ON-TARGET COVERAGE ENRICHMENT
10	230	0.58	297
11	80	0.23	347

DISCUSSION

Significant time savings and increased flexibility with Monarch HMW DNA Extraction Kit

Until recently, Cas9 enrichment workflows have been described using phenol/chloroform extracted HMW DNA when working with animal tissues. The drawbacks of this extraction approach are numerous, including significant handling, hazardous chemicals, and excessive time required for dissolving. Because this method requires several days to obtain usable DNA, it is not conducive to rapid transgene/strain analysis and limits the flexibility needed for troubleshooting or tweaking of parameters.

The new Monarch HMW DNA extraction workflow provided DNA with high yields, high purity, and high DNA integrity which was ready to use in only a few hours. The Monarch-extracted DNA performed well in Cas9 sequencing and resulted in a significant time savings in the Cas9 enrichment sequencing workflow of up to 3 days (Figure 8).

The Monarch workflow also enables tunable fragment length generation by having the user change the agitation speed of the thermal mixer during lysis, empowering the user to adjust the size of the DNA to optimize conditions for the relevant downstream application. This property proved useful, as it allowed for troubleshooting of experiments that failed with highly viscous ultra-high molecular weight DNA which had been isolated after lysis at a low agitation speed (500 rpm). A repeat of the same Cas9 sequencing experiment with less viscous samples that were purified with 2000 rpm agitation speed during lysis, led to sufficient coverage of the region of interest (Samples 6 and 7, described more in detail below).

Monarch extraction outperforms phenol/chloroform in most metrics

DNA yield

The Monarch protocols provide good yields for both mouse tissue types investigated. Yields per mg tissue for liver samples were similar for both extraction methods at around 1.5 µg/mg tissue. Actual yields with phenol may be significantly lower if Qubit values are taken as standard. For kidney, yields per mg were high with both methods, with phenol giving the largest overall yield. Troubleshooting of the lower Monarch yields

FIGURE 4: **Typical pore usage of Cas9 targeted run**

Screenshots taken from the GridION X5 software during the sequencing of Sample 10 and Sample 11. After 24 hours sequencing of Sample 10 (A), the flow cell was nuclease washed, primed, and loaded with Sample 11 and ran for a further 24 hours (B).

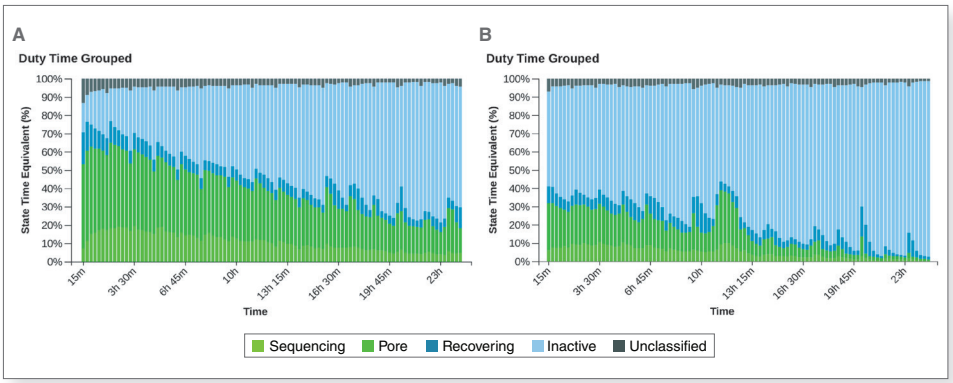


TABLE 7: **Cas9 capture sequencing MinION and GridION run metrics**

Run metrics from the multiple 24-hour targeted sequencing runs are summarized. Samples 1, 10 and 11 were run on Oxford Nanopore Technologies GridION platform, while all others were run on a MinION.

MOUSE LINE	MEAN READ LENGTH (BASES)	MEAN READ QUALITY	MEDIAN READ LENGTH (BASES)	MEDIAN READ QUALITY	NUMBER OF READS	READ LENGTH N50 (BASES)	TOTAL BASES
1	15,649	11.9	5,949	12.8	67,863	41,296	1,062,009,255
4, 5	16,282	9.9	7,152	10.6	49,790	38,666	810,697,934
6	7,713	11.0	1,391	11.2	14,306	37,682	110,347,242
7	4,728	10.5	969	10.6	17,017	23,816	80,452,590
8, 9, 12, 14	12,588	10.7	6,392	10.9	57,083	26,806	718,531,162
10	14,552	12.4	6,596	12.7	111,898	32,880	1,628,300,195
11	10,188	11.9	3,983	12.1	64,636	26,210	658,525,432

FIGURE 5: **Comparison of MX1 in Mouse Lines C57BL/6J and CAST/EiJ**

(A) Sample 10: 230X coverage depth over a 19 kb region of C57BL/6J spanning the entire MX1 locus aligned to MM10. (B) Sample 11: 80X coverage of a 22.5 kb region from CAST/EiJ spanning the MX1 locus and identifying the 3.5 kb not present in C57BL/6J aligned to MM10 (C) Reads covering MX1 C57BL/6J when aligned to CAST/EiJ

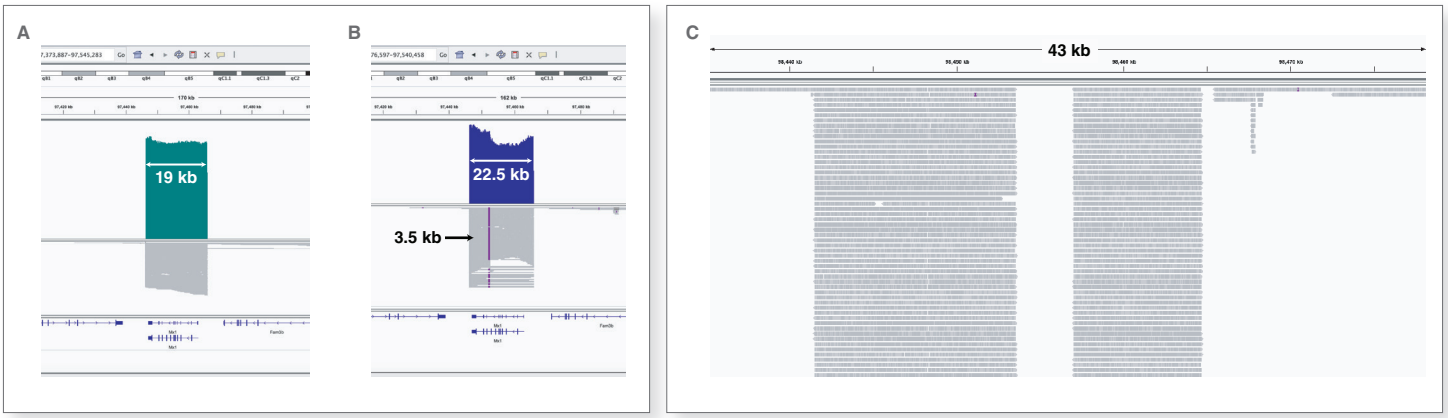
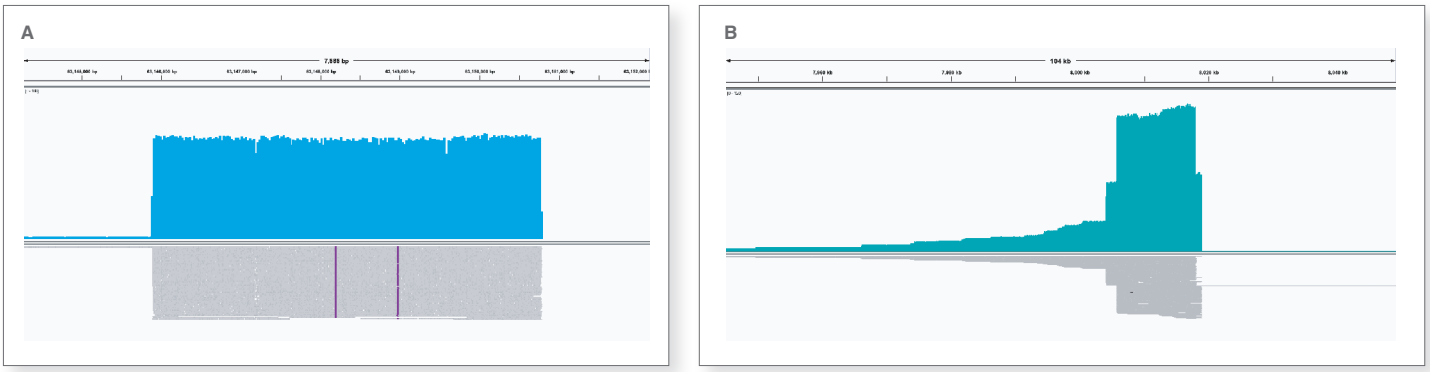


FIGURE 6: **Coverage depth and associated reads validating samples with targeted mutations**

(A) Sample 8: homozygous 5 kb region covering a floxed exon. Sequencing generated 70X converge over the region of interest. (B) Sample 1: 13 kb region of interest for the validation of a targeted mutation. Capture sequencing generated a coverage depth of 100X over region of interest. Reads from 13 to 95 kb in length were generated.



observed with 4 kidney samples prepared with low agitation speeds during lysis revealed that for such samples, increasing the length of the binding step in the vertical rotating mixer is recommended in the protocol. This extended binding time resulted in higher yields in the next set of DNA extractions from kidney tissue. Although phenol preps by nature offer more flexibility regarding the maximal sample input amounts, the total yields obtained using the Monarch kit were good and sufficient for several experiments. Moreover, recent Cas9 sequencing experiments performed with Monarch-purified DNA indicate that high coverage in Cas9 experiments can be achieved with significantly lower DNA input amounts (data not shown).

DNA quality

Samples isolated using the Monarch kit were higher purity than those isolated with phenol/chloroform, particularly with liver samples (A_{260}/A_{230} with phenol ranged 1.1-1.3). Also, the rapid DNA integrity analysis on the TapeStation indicated that DNA isolated using the Monarch kit had a higher proportion of DNA > 50 kb when compared with that isolated by phenol extraction. Consequently, using Monarch as an alternative to phenol extraction is not only a faster approach, but may also result in better performance in long read sequencing applications.

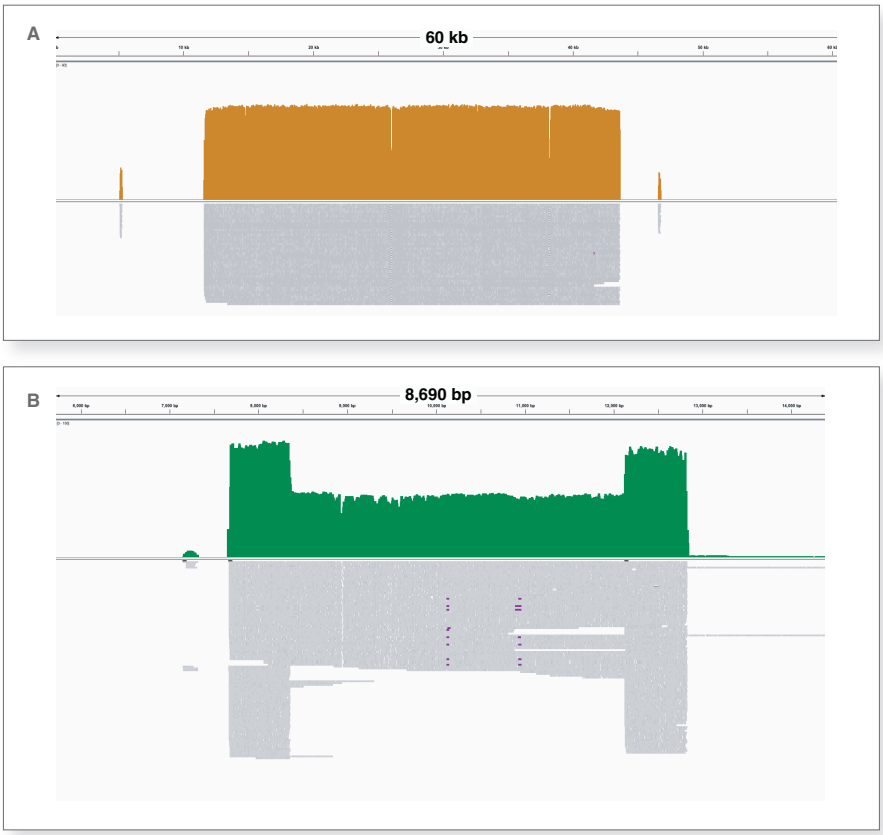
Targeted sequencing results

Cas9-enrichment for comparison of wild-derived and common inbred laboratory strains

It has been known for some time that most inbred laboratory mice do not possess genetically-mediated resistance to Myxoviruses (6,7) . The interferon-inducible protein responsible for this resistance is encoded by MX1 located on Chromosome 16. Several inbred mouse strains have acquired a large deletion or nonsense mutations in MX1 due to a founder effect resulting in a non-functional protein (8). The absence of a functional Mx1 protein in C57BL/6J mice has been linked to an increased susceptibility to Influenza A virus infection. In contrast, recently wild-derived strains, including CAST/EiJ contain a functional MX1 allele that renders them highly resilient to viral infections. The high coverage depth generated by Cas9 targeted sequencing of the entire MX1 locus identified the large 3.5 kb deletion present in the common laboratory mouse C57BL/6J (Figure 5, Samples 10 and 11).

FIGURE 7: **Analysis of 30 kb homozygous transgene insertion and 5 kb heterozygous transgene insertions**

(A) Sample 5: Homozygous 30 kb insert; coverage across the region of interest was 44X. (B) Sample 4: a mouse line with a 5 kb heterozygous insertion, coverage depth of 92x. Note: mean coverage depth does not include endogenous chromosomal reads located at the 5´ and 3´ ends of the transgenic insertion.



Analysis of transgene-targeted mutagenesis sites

Figure 6A, Sample 8 illustrates the power of this technology for validating mouse models. Targeted mutagenesis was carried out to introduce loxP sites flanking exon 4 of our gene of interest. To assess the integrity of the loxP sites and the surrounding chromosomal context, sgRNAs were designed up and downstream to excise a 5 kb fragment. Cas9 enrichment was carried out, producing 70X coverage of the region of interest. Sequence alignment revealed the presence of two unexpectedly large inserts – one 180 bp and another 80 bp (Figure 6A). These larger-than-expected insertions are suspected to be a result of plasmid DNA integration along with the loxP sequence.

FIGURE 8: **Comparing the Cas9 sequencing workflow duration for phenol/chloroform and Monarch HMW DNA extraction**

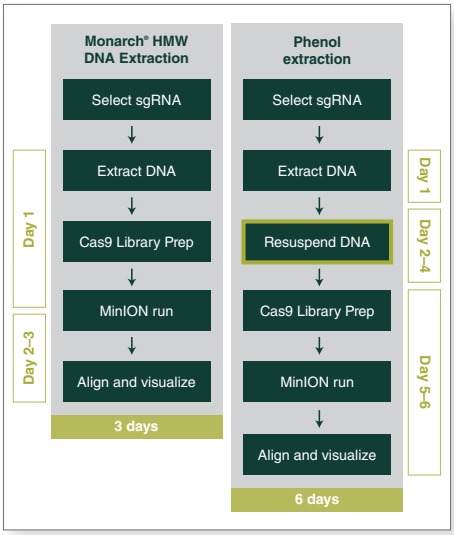


Figure 6B, Sample 1, depicts the targeted sequencing of a 13 kb region at the terminal end of the gene of interest, covering exons 25–29. The model in question was subject to targeted mutagenesis within this region. Cas9 capture sequencing was employed to assess the success of the genetic modification. The subsequent sequencing yielded a mean of 100X coverage over our target region whilst generating reads 13-95 kb in length. Larger insertions than expected were revealed, indicating the requirement for further investigation into this region. The Cas9 targeted sequencing approach saved considerable time as traditional Sanger-based methods did not yield any insights into this particular locus condition.

Cas9-enrichment is a powerful tool for assessment of transgene insertion sites

The efficiency of this method is highlighted in Figure 7, showing analysis of Sample 5 (homozygous 30 kb insert) and Sample 4 (heterozygous 5 kb insert). Sample 5 contains a large insert and delivered 44X coverage over the region of interest. In Samples 4 and 9, the Cas9 enrichment approach was used to check the integrity of the insertion site as well as the transgene sequence. The sgRNAs were designed 1 kb up and downstream of the transgene, and the enrichment approach worked well with a 92X and 74X coverage, respectively. A slight overrepresentation of the shorter allele not containing the transgene was observed.

Sample 6 and 7 (both heterozygous 5 kb inserts) proved to be difficult targets; even with the Cas9 enrichment approach, it was challenging to generate reads. Monarch HMW DNA extraction was repeated with larger input amounts and lysis was carried out at maximal agitation speed (2000 rpm) to reduce the size of the HMW DNA fragments and the accompanying viscosity. This modification enabled 20X/23X coverage of the region of interest, respectively, sufficient for the application. This example provides a compelling case of how having access to a rapid extraction method enables troubleshooting experiments without significant time loss.

CONCLUSION

Overall, the Cas9 enrichment approach is a powerful tool for interrogation of genomic loci. Having a rapid high-quality method like Monarch for HMW DNA extraction enables a significant reduction in the workflow time and facilitates troubleshooting efforts without adding several days of work. Looking forward, we aim to develop Cas9 enrichment-based transgene screening approaches that do not require the sacrifice of the animals, as they currently do when HMW DNA needs to be isolated from organ tissue. Future tests will, therefore, focus on using ear punch or tail clip tissue as input material for HMW DNA extraction and optimizing Cas9 enrichment sequencing for low input samples.

IN SUMMARY

- The Monarch HMW DNA Extraction Kit enabled a reduction of the total Cas9 enrichment workflow time by 3 days. In addition to the significantly faster DNA extraction workflow, better solubility of the Monarch purified HMW DNA resulted in significant time savings.
- The HMW DNA isolated with the Monarch kit consistently was of high yield, superior quality and longer fragment length; thus, this approach proved to be a powerful alternative to phenol chloroform extractions.
- Tunable fragment length of the Monarch-isolated HMW DNA is a valuable feature when troubleshooting targeted sequencing results.
- Coupling the Monarch HMW DNA extraction procedure with an optimized Cas9 no-amp enrichment library prep protocol has resulted in the successful establishment of a highly efficient standard procedure for the analysis of transgene insertion sites in mice by Genome Technologies at The Jackson Laboratory.

References

1. Goodwin, L.O. et al. (2019) *Genome Research.* https://doi.org/10.1101/gr.233866.117.
2. Gilpatrick, T. et al. (2020) *Nature Biotechnology* 38, 433–438.
3. Oxford Nanopore Technologies (2020). "Cas9 Targeted Sequencing Before Start Checklist Cas9 Targeted Sequencing," 1–6.
4. Gong, L. et al. (2019) *Journal of Visualized Experiments: JoVE.* https://doi.org/10.3791/58954.
5. Pacific Biosciences. 2019 "Reference Guide – Designing CRISPR-Cas9 RNA Oligonucleotides for the No-Amp Targeted Sequencing Procedure." *Pacific Biosciences* 01 (101): 1–10.
6. Lindenmann, J. et al. (1962) *The Resistance of A2G Mice to Myxoviruses.*
7. Tumpey, T.M., et al. (2007) *Journal of Virology* 81, 10818–10821.
8. Staeheli, P, et al. (1988). *Molecular and Cellular Biology.* 8, 4518-4523.