1: Library Construction Protocol

Genomic DNA is first shotgun sheared to a nominal 1 kbp fragment length, using commercially available ultrasonic or hydraulic shearing systems. The Polonator protocols have been tested at the nominal 1kbp fragment length, although users are free to try a wide range of alternative fragment lengths (~300 bp – 10 kbp), as their application dictates. Note, however, that since the paired tag protocol only reads 26 bases per fragment, more genomic DNA will be required at longer fragment lengths. The shearing process typically leaves overhangs at each end. The fragments are purified, blunted (end repaired to remove any overhangs), A-tailed at the 3’ ends, and then size selected.

DNA shotgun fragment, after shearing, blunting and A-tailing:

```
      ????????????????????????.....????????????????????A   (length ~1K bp)
      ||||||||||||||||||||||||.....||||||||||||||||||||||
     A????????????????????????.....????????????????????
```

Complementary strands of a synthetic 30-mer with T-tailed 3’ ends and phosphorylated 5’ ends are annealed:

```
T30:   GTTGGACGTACGGCCGCCTTGGCCTCCGACT
       ||||||||||||||||||||||||||||||
       TCAACCTGCATGCCGGCGGAACCGGAGGCTG
```

The two ends of the 1 kb long, A-tailed genomic DNA are then “circularized” (ligated, with conditions favoring the formation of monomeric recombinant circles) to the T-tailed ends of the synthetic 30-mer. After ligation, the center synthetic segment is 32 bp long. The circularization process is relatively inefficient, but subsequent amplification more than makes up for its low yield.

A character-based zoom in on the relevant part of the resulting plasmid-like circle is shown here:

```
.....?????????????????????AGTTGGACGTACGGCCGCCTTGGCCTCCGACT?????????????????????.....
.....|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||.....
.....?????????????????????TCAACCTGCATGCCGGCGGAACCGGAGGCTGA?????????????????????.....
```

And a more ambitious graphical rendering is here:

Any residual noncircularized material is eliminated by exonucleosis with exonuclease I and III.

The circularized genomic DNA:32-mer is then amplified by hyper-branched rolling-circle amplification, using a GE TempliPhi™ kit. This amplifies the number of circular molecules ~10^7-10^8 fold, while not amplifying any linear material from the preceding steps.

The amplified circular material is then digested with the restriction endonuclease Mme I, which binds to its recognition sequence at each end of the synthetic 32-mer, and cuts 17 to 18 bp into each end of the genomic DNA (the enzymatic cutting has approximately equal odds of cutting 17 or 18 bp into the genomic DNA). This results in a 66 to 68 bp component, consisting of the synthetic 32-mer with a 17 to 18 bp genomic tag on each end. As a result of the Mme I digestion, there is also a two-nucleotide overhang at each 3' end:

```
   ??????????????????AGTTGGACGTACGGCCGCCTTGGCCTCCGACT????????????????????
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 ??????????????????TCAACCTGCATGCCGGCGGAACCGGAGGCTGA??????????????????
```

A gel separation is then used to separate the above population of 66 to 68 bp oligos from the much longer (~965 bp) remains of the original genomic DNA (which is discarded).

The 66 – 68 bp long oligo is then blunted to remove the two-nucleotide 3' overhangs left over from the Mme I digestion:

```
   ??????????????????AGTTGGACGTACGGCCGCCTTGGCCTCCGACT??????????????????
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 ??????????????????TCAACCTGCATGCCGGCGGAACCGGAGGCTGA??????????????????
```

The 5' ends of the above 66 – 68 bp oligo are phosphorylated.

Two complementary strands of a 43 bp forward primer (FDV-T and FDV-B) are annealed to produce:

```
FDV:  AACCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT
      |||||||||||||||||||||||||||||||||||||||||||
      TTGGTGATGCGGAGGCGAAAGGAGAGATACCCGTCAGCCACTA
```

Two complementary strands of a 25 bp reverse primer (RDV-T and RDV-B) are then annealed to produce:

```
RDV:  AGAGAATGAGGAACCCGGGGCAGTT
      |||||||||||||||||||||||||
      TCTCTTACTCCTTGGGCCCCGTCAA
```

Neither the forward or reverse primers are 5' phosphorylated. The forward and reverse primers are then ligated to their respective ends of the 66 to 68 bp component, producing a 134 – 136 bp long oligo which consists of the 43 bp forward primer, the first 17 to 18 bp genomic tag, the synthetic 32-mer, the second genomic tag, and finally the 25 bp reverse primer. Since each primer's 5' ends aren't phosphorylated on the end that ligates to the genomic tags, this leaves a nick (a missing phosphor-diester bond) at these two points. E.coli DNA polymerase I and a dNTP mix are then used to translate this nick off each end of the oligo.

Here is the resulting 134 – 136 bp long oligo, referred to as “template DNA”:

```
AACCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT????????????????????AGTTGGACGTACGGCCGCCTTGGCCTCCGACT????????????????????AGAGAATGAGGAACCCGGGGCAGTT
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
TTGGTGATGCGGAGGCGAAAGGAGAGATACCCGTCAGCCACTA????????????????????TCAACCTGCATGCCGGCGGAACCGGAGGCTGA????????????????????TCTCTTACTCCTTGGGCCCCGTCAA
```

The template DNA is then amplified using PCR, which also suppresses FDV-genomic tag-FDV and RDV-genomic tag-RDV ligates.
With only 12 cycles this PCR step, while useful, introduces very little bias.

2. Emulsion PCR Protocol

The components of a solid phase (bead-based) PCR emulsion are then mixed, which include beads, oil, template DNA from the previous library construction protocol, a mix of dNTPs, buffers, Taq polymerase, and free reverse primer. The beads are monosized polystyrene spheres which are 1 um in diameter; they are superparamagnetic to allow magnetic separation, and are coated with streptavidan. The following forward primer has been preloaded to the bead through a streptavidan- dual biotin coupling:

Bead-Streptavidin-Dual_biotin-CCACTACGCCTCCGCTTTCCTCTCTATGG

The free reverse primer is as follows:

TCTTACTCCTTGGGCCCCGTC

Dilution of the template DNA and bead suspension is employed such that once emulsified, the majority of the aqueous droplets will contain only one molecule of template DNA, along with only one primer-loaded polystyrene bead. Droplets with two molecules of template DNA will generate erroneous reads and be eliminated from the output data by the image processing software.

Vortexing is then used to create the emulsion, producing aqueous droplets of ~ 10 micron diameter.

Next, PCR is initiated. Each emulsion droplet functions as an independent PCR reservoir. In each, the single double stranded molecule of template DNA is thermally denatured at 94ºC for 15 seconds to yield the following two single stranded molecules:

AACCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT????????????????????AGTTGGACGTACGGCCGCCTTGGCCTCCGACT????????????????????AGAGAATGAGGAACCCGGGGCAGTT

and:

TTGGTGATGCGGAGGCGAAAGGAGAGATACCCGTCAGCCACTA????????????????????TCAACCTGCATGCCGGCGGAACCGGAGGCTCA????????????????????TCTCTTACTCCTTGGGCCCCGTCAA

The temperature is then lowered to 57ºC for 30 seconds to allow the lower strand to anneal to the bead-bound forward ePCR primer:

```
biotin-CCACTACGCCTCCGCTTTCCTCTCTATGG
       ||||||||||||||||||||||||||||
     TTGGTGATGCGGAGGCGAAAGGAGAGATACCCGTCAGCCACTA????????????????????TCAACCTGCATGCCGGCGGAACCGGAGGCTCA????????????????????TCTCTTACTCCTTGGGCCCCGTCAA
```

and the upper strand to anneal to the free reverse ePCR primer:

```
AACCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT????????????????????AGTTGGACGTACGGCCGCCTTGGCCTCCGACT????????????????????AGAGAATGAGGAACCCGGGGCAGTT
                                                                                                                    ||||||||||||||||||||
                                                                                                                    TCTTACTCCTTGGGCCCCGTC
```

The temperature is then raised to 70ºC for 75 seconds, allowing the Taq polymerase to complete dNTP incorporation. We now have two complete double-stranded DNAs, one bound to the bead, and the other free:

Bead-bound, double-stranded DNA:

```
biotin-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT????????????????????AGTTGGACGTACGGCCGCCTTGGCCTCCGAGT????????????????????AGAGAATGAGGAACCCGGGGCAGTT
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     TTGGTGATGCGGAGGCGAAAGGAGAGATACCCGTCAGCCACTA????????????????????TCAACCTGCATGCCGGCGGAACCGGAGGCTCA????????????????????TCTCTTACTCCTTGGGCCCCGTCAA
```

Free double stranded DNA:

```
AACCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT????????????????????AGTTGGACGTACGGCCGCCTTGGCCTCCGACT????????????????????AGAGAATGAGGAACCCGGGGCAGTT
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
TTGGTGATGCGGAGGCGAAAGGAGAGATACCCGTCAGCCACTA????????????????????TCAACCTGCATGCCGGCGGAACCGGAGGCTGA????????????????????TCTCTTACTCCTTGGGCCCCGTC
```

120 of these PCR cycles are performed, vastly amplifying the number of bead-bound primer – template DNA molecules.

3. Emulsion Breaking Protocol:

This one is pretty brief. The emulsion is broken using isopropanol, and the beads are separated from the oil and free DNA using a series of centrifuge and magnetic separation steps. Once separated, an NaOH solution is used to denature and remove the lower strand from the bead-bound template DNA.

This leaves us with ~400,000 of the following single-stranded template DNA molecules, 130 – 132 bp long, bound to each bead:

```
bead- dual biotin-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT????????????????????AGTTGGACGTACGGCCGCCTTGGCCTCCGAGT????????????????????AGAGAATGAGGAACCCGGGGCAG
```

The genomic DNA nearest the bead is referred to as the proximal tag “A”, while the other is referred to as the distal tag “B”.

4. Enrichment Protocol:

Some of the emulsion PCR droplets had no template molecule, and are referred to as “unamplified”. We would like to get rid of these, and so we “enrich” the amplified beads.

We begin by binding a synthetic 69-mer to a solution of larger (3 um diameter) “capture beads”, which are not superparamagnetic (to facilitate their subsequent separation from the template DNA beads).

Here is the 69-mer we bind to the larger capture beads:

5'-biotinTEG-CGTACCCCGCTTGGTCTTTCTCCCGTACCCCGCTTGGTCTTTCTCCCTGCCCCGGGTTCCTCATTCTCT

We mix the ePCR beads with a solution of these “capture beads” with their bound oligo. The last 23 bases of the latter hybridize with their complementary sequence at the 3’ end of the single-stranded, bead-bound template DNA. Even with a 12 point font and an 11” x 17” page format in landscape mode, this complex doesn’t fit, so we can’t spell it out, as it were, but we suspect you’re still with us.

This template DNA bead-capture bead complex has a lower effective density than the unamplified 1 um beads, which do not complex with the capture beads because their unextended forward primers cannot hybridize with the capture bead oligo. We then perform a centrifugal separation. The low density beads and their primers are subsequently separated from the amplified beads by melting with NaOH, and magnetic separation is used to harvest the enriched, amplified beads.

5. Bridging and Capping Protocol:

Two bridge oligos (Bridge 1: GTGAGCTTCGTCTGCCCCGGGTTC Bridge 2: GTGAGCTTCGTCCATAGAGAGGAAAGCG,) are now added; Bridge 1 hybridizes to and overhangs the 3’ end of template DNA, while Bridge 2 hybridizes to and overhangs the 3’ end of unextended forward primers. Here is the template DNA with Bridge 1:

```
biotin-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT??????????????????AGTTGGACGTACGGCCGCCTTGGCCTCCGAGT??????????????????AGAGAATGAGGAACCCGGGGCAG
                                                                                                                    ||||||||||||
                                                                                                                    CTTGGGCCCCGTCTGCTTCGAGTG
```

Here Bridge 2 is shown hybridized to an unextended forward primer:

```
biotin-CCACTACGCCTCCGCTTTCCTCTCTATGG
                  ||||||||||||||||
                  GCGAAAGGAGAGATACCTGCTTCGAGTG
```

Three capping oligos are then introduced, along with ligase: Cap 1:CGAAGCTCAC-Amn, Cap 2: ACGAAGCTCAC-Amn, and Cap 3: NNNNNNNNN-Amine. Cap 1 and Cap 2 hybridize to the overhanging bridge oligos, and ligate to the template DNA and the forward primer. The previous ePCR reaction yields two fractions of product: with A tails added on the 3' ends, and without. The only difference between Cap 1 and Cap 2 is the leading "A", to permit the capping of both A-tailed (Cap 1) and non-A-tailed species (Cap 2). Since the melt temperature is quite low, Cap 1 and Cap 2 do not compete with each other; only the proper length cap will both hybridize and ligate.

Here are the final two bead components after capping (an A-tailed molecule of template DNA is shown):

```
biotin-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT??????????????????AGTTGGACGTACGGCCGCCTTGGCCTCCGAGT??????????????????AGAGAATGAGGAACCCGGGGCAGACGAAGCTCAC-Amn
                                                                                                                        ||||||||||||||||||||||||
                                                                                                                        CTTGGGCCCCGTCTGCTTCGAGTG
```

Depending on the length of the two genomic tags (either 17 or 18 bp), the above oligo will have a length of between 141 and 143 bp.

```
biotin-CCACTACGCCTCCGCTTTCCTCTCTATGGACGAAGCTCAC-Amn
                  ||||||||||||||||||||||||||||
                  GCGAAAGGAGAGATACCTGCTTCGAGTG
```

Cap 3, the third, degenerate component, caps other species that may be present, including incomplete extension products. It is added after the other two caps to avoid undesired ligation with template DNA.

After the caps are added, the bridges are melted off with NaOH, and Bridge 1 is then re-hybridized (the net result of these two steps is that the desired template DNA is double-stranded at the far end from the bead, while the unextended forward primer will be single stranded (so that it can be digested with exonuclease I at the beginning of the sequencing protocol).

The capped beads are then magnetically separated and washed, and are now ready to be arrayed.

6. Aminosilanation Protocol:

The OpenWetWare wiki shows the aminofunctionalization of the coverslip glass as a separate protocol, but treated coverslips are now provided as a part of each flow cell kit. Users can easily pipette individual libraries into the 18 wells of each flow cell and bind them as described at the end of the Bead Capping Protocol at the OpenWetWare wiki. The amine group terminating all three capping oligos couples the beads to the aminosilinated glass via a BS3 covalent linker.

7. Sequencing Protocol:

All functions of this protocol are implemented by the Polonator. From one perspective, the Polonator's mission is simply the automation of the seventh protocol.

This protocol begins with the digestion of single-stranded, unextended, bead-bound forward primers, using Exonuclease I. This is then denatured with 6M Guanidine HCL, and the bridging oligos are melted off with 0.1M NaOH.

A white light imaging cycle determines the position of all two billion beads (one billion per flowcell).

Sequencing uses anchor primers A1 to A4, as well as N1 and N3; anchors A1 and A3 are 5' phosphorylated:

Anchor A1: 5' Phos-ATCACCGACTGCCCA

Anchor A2: 5' CGGCCGTACGUCCAACT

Anchor A3: 5' Phos-AGTCGGAGGCCAAGG

Anchor A4: 5' CCCGGGTTCCTCATTCTCT

Anchors N1 and N3 are nearly identical to anchors A1 and A3, but add one degenerate base, which overhangs their respective tags by one base, and allows one base further into the tag to be read. Anchors N1 and N3 are used only once per run, while anchors A1, A2, A3, and A4 are each used six times.

Anchor N1: 5' Phos-NATCACCGACTGCCCA

Anchor N3: 5' Phos-NAGTCGGAGGCCAAGG

While they are actually used one at a time, here's how anchors A1 to A4 fit onto the template DNA:

```
                                        Proximal Tag "A"                                                    Distal Tag "B"

biotin-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT????????????????????AGTTGGACGTACGGCCGCCTTGGCCTCCGAGT????????????????????AGAGAATGAGGAACCCGGGGCAGACGAAGCTCAC-Amn
                                  ||||||||||||||||                    ||||||||||||||||||||||||||||||||||                    ||||||||||||||||||||
                                  ACCCGTCAGCCACTA                      TCAACCUGCATGCCGGCGGAACCGGAGGCTGA                      TCTCTTACTCCTTGGGCCC
                                     Anchor A1                            Anchor A2       Anchor A3                            Anchor A4
```

And here's the template DNA with anchors N1 and N3 hybridized, along with A3 and A4 (again, only one anchor is actually hybridized at a time; they are all shown together simply for visual purposes):

```
                                        Proximal Tag "A"                                                    Distal Tag "B"

biotin-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT????????????????????AGTTGGACGTACGGCCGCCTTGGCCTCCGAGT????????????????????AGAGAATGAGGAACCCGGGGCAGACGAAGCTCAC-Amn
                                  ||||||||||||||||                    ||||||||||||||||||||||||||||||||||                    ||||||||||||||||||||
                                  ACCCGTCAGCCACTAN                     TCAACCUGCATGCCGGCGGAACCGGAGGCTGAN                     TCTCTTACTCCTTGGGCCC
                                     Anchor N1                            Anchor A2       Anchor N3                            Anchor A4
```

Reading is performed by hybridizing an anchor primer to the template DNA. A mixture of DNA ligase, ligation buffer, and a population of 4^9 (262,144) query nonamers is then introduced. Each nonamer is terminated with one of four fluorophores, with a 1:1 mapping between the fluorophore and the base at the query position. The four current fluorophores are 6FAM, Ty563, Tex 613, and Ty665; the respective visual color of their excitation wavelengths are blue, green, yellow, and red, respectively. If the query position in the nonamer is complementary to the base in the genomic tag at that location, then that nonamer ligates to the anchor primer and hybridizes with the tag. A four-color imaging cycle then reads the base at that query location for all the beads in the flow cell.

The mapping of the four fluors to the four bases is as follows:

Fluor	Nonamer Query Base	Genomic Tag Base Read
6FAM	G	C
Ty563	A	T
Tex613	C	G
Ty665	T	A

For both the proximal tag “A” and the distal tag “B”, seven bases are read on the side closest to the bead (designated as the “minus” side), and six bases are read from the side away from the bead (the “plus” side). Here’s how the bases are numbered for the proximal tag “A” (the same pattern is repeated for the distal tag “B”):

```
          “minus”     “plus”
          1234567     654321
...TCGGTGAT???????????????????AGTTGG...
```

On the proximal tag “A”, base read minus7 uses anchor N1; base reads minus 6 through minus 1 use anchor A1; base reads plus 6 through plus 1 use anchor A2. On the distal tag “B”, base read minus 7 uses anchor N3; base reads minus 6 through minus 1 use anchor A3; base reads plus 6 through plus 1 use anchor A4. Notice that 4 – 5 bases at the center of each tag go unread.

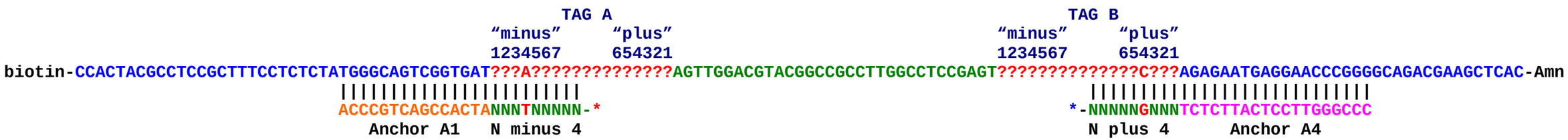
Here’s what a pool of nonamers look like for interrogating base position minus 4 for anchor primers A1 and A3:

```
  5' 6FAM-NNNNNGNNN
 5' Ty563-NNNNNANNN
5' Tex613-NNNNNCNNN
 5' Ty665-NNNNNTNNN
```

Note that the 5’ ends of these nonamers are not phosphorylated. Here is the same pool of nonamers for interrogating base position plus 4 for anchors A2 and A4; these nonamers will ligate on their right side, and their 5’ ends are accordingly phosphorylated:

```
5' Phos-NNNGNNNNN-6FAM
5' Phos-NNNANNNNN-Ty563
5' Phos-NNNCNNNNN-Tex613
5' Phos-NNNTNNNNN-Ty665
```

Here is the bead bound template DNA with anchors A1 and A4, with the above nonamers that interrogate base positions AM4 (Tag A, base read minus 4) and BP4 (Tag B, base read plus 4). Note that only one anchor and nonamer are hybridized and ligated at a time; the two are shown at the same time simply for visual purposes.



The "N"s refer to all four possible bases; recall that each nonamer-fluor combo is actually present in 4^9 (262,144) species, with all possible combinations of bases present (again, for a given query position, the base at that position uniquely defines the nonamer's fluorophore). There are six different nonamer pools for the minus side base reads (referred to as N-1 to N-6), and six different nonamer pools for the plus side base reads (N+1 to N+6). The nonamer N-6 is used twice at both the proximal and distal tags (see below).

Reading begins by hybridizing anchor N1, and then performing a ligation with nonamer pool N-6; since the anchor N1 overhangs the tag by one base, this actually reads base position minus 7. After each ligation, a four-color imaging scan of the entire flow cell is performed. 6M Guanidine is introduced to denature residual enzyme, and the anchor-nonamer ligate is melted off with NaOH. Anchor A1 is then hybridized, and nonamer pool N-6 is used again, this time to read base position minus 6. After denaturing and melting, anchor A1 is hybridized again, this time with nonamer pool N-5, reading base position minus 5. This hybridize-ligate--four color image-denature-melt process is then repeated with anchor A1 four more times, each time reading one base closer to the anchor. After reading base position minus 1 (directly adjacent to the anchor), the process is repeated on the plus (right-hand) side of the proximal tag, using anchor A2 six times, with nonamer pools N+6 to N+1. A total of thirteen reads are performed on the proximal tag (seven on the minus side, and six on the plus side; 4 to 5 bases in the center of the tag remain unread). The above process is repeated for the distal tag, using anchors N3, A3, and A4. With 13 bases read from each of the two tags, the total number of bases read is 26 per readable bead. With ~ 2 billion beads in the two flow cells, and 26 bases per bead, a total run output of 52 G base pairs is in principal possible. At present, output for a sequencing run is about 10 G base pairs, since only one in six beads currently provides a clear read. Efforts are underway to raise the percentage of readable beads.

The duration of a sequencing run is ~ 80 hours; 90 minutes of biochemistry and 90 minutes of imaging for each of 26 bases. In all, ~5.7 million, one-megapixel bead images will be acquired and processed in real time.

Kevin McCarthy
Chief Technology Officer
Dover Motion Systems
603-870-5611
Kevin.mccarthy@danahermotion.com
© Dover Motion Systems 2008