Sequencing single human cells and bacterial DNA using PicoPLEX DNA-seq at low coverage for aneuploidy, CNV, and genotyping applications

J Langmore, E Kamberov, T Tesmer, S. Yerramilli, M. Carey, M. Carroll. Rubicon Genomics, Ann Arbor, MI

www.rubicongenomics.com • +1 734.677.4845

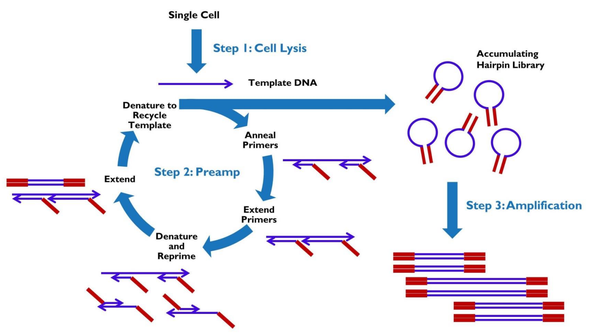
Abstract

Advances in whole genome amplification (WGA) and microarrays have shown >30% improvement in the success of in vitro fertilization (IVF) in conjunction with PGS to prevent implantation of aneuploidy embryos. NGS will replace array-based and perhaps PCR-based PGS/PGD in the near future if a) the results are equivalent or superior to current methods with respect to accuracy and resolution of aneuploidy and CNV detection; b) can be reported less than 16h after receipt of the embryo biopsies; and c) cost less. This presentation will address those three criteria for single-blastomere and multiple-cell trophectoderm testing.

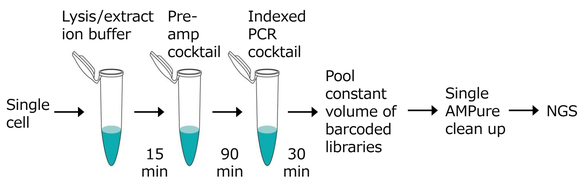
We used the Rubicon PicoPLEX™ DNA-seq single-cell NGS library prep kit to detect chromosomal and segmental aneuploidies with the 10 – 20 Mb resolution as provided by the 24sure BAC arrays by BlueGnome© (an Illumina© company). Equivalency was achieved on the Illumina MiSeq™ with 100,000 – 200,000 total reads in less than 8 hours (50 b genomic read and two 8b indexing reads). The PicoPLEX single cell prep requires 3h for cell lysis, DNA extraction, library synthesis, amplification, pooling and AMPure™ clean up. Only three additions of reagents (and no intermediate clean ups) are used to create the amplified libraries, and no quantification of libraries necessary before pooling. Time to result is no more than 12 hours to test 96 – 192 samples on the MiSeq.

NGS of single human cells show base coverage >30% at a sequencing depth of 3X. NGS of E. coli shows 90% bases are covered at a sequencing depth of 20X. NGS of 5-cell equivalent trophetoderm samples show that terminal translocations as small as 6 Mb can be detected at this low coverage. Typical 100 b paired-end performance on the MiSeq v2 is 0.9 – 1.1 M clusters/mm2; 85-95% PF; and 97% human reads with <0.3% mismatch to hg19 and <1% chimeric inserts. Experience shows that the same NGS libraries can be used for PGD, to detect familial single-gene disorders using SNP-based or STR-based locus-specific linkage assays. Thus, low-coverage NGS meets the requirements for IVF testing.

PicoPLEX DNA-seq Thermal Cycling Quasi-Random Primed Library Chemistry



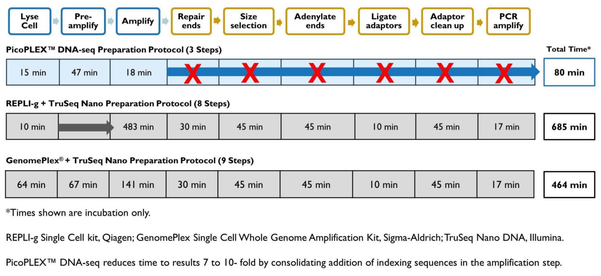
PicoPLEX DNA-seq Workflow



Characteristics of PicoPLEX

- Designed and optimized for rapid, reproducible CNV detection and signature sequencing of infectious diseases
- "No Strand Left Behind™ lysis and extraction
- Non-complementary quasi-random primers with undetectable primer-dimers
- "Multi-pass" thermo-cycling library synthesis for reproducibility
- High systematic bias for deep sequencing at low coverage
- 48 Dual-Read bar codes
- All reagents are thermostable
- No quantification before sample pooling or sequencing
- Single AMPure clean up before sequencing

PicoPLEX DNA-seq Workflow Advantage



*Times shown are incubation only.

REPLI-g Single Cell kit, Qiagen; GenomePlex Single Cell Whole Genome Amplification Kit, Sigma-Aldrich; TruSeq Nano DNA, Illumina.

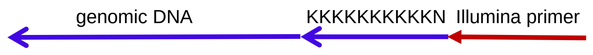
PicoPLEX™ DNA-seq reduces time to results 7 to 10- fold by consolidating addition of indexing sequences in the amplification step.

PCR and Array Genotyping of PicoPLEX Single-Cell Libraries (Dagan Wells, University of Oxford)

SNP GENOTYPING METHOD	SINGLE-CELL AMPLIFICATION SUCCESS RATE	SNP CALL RATES	LOSS OF HETEROZYGOSITY
PCR	95%	>95%	<10%
Illumina SNP array	95%	50% - 60%	7% - 12%

Single-Cell NGS Using PicoPLEX DNA-seq

- Standard Illumina sequencing and indexing primers
- Broad insert size distribution (100 – 2000 bp)
- Reproducibility makes quantification unnecessary
- Quasi-random primers pose diversity problems at beginning of read, however three solutions have been tested successfully:
  - Do nothing
  - Add 5% phiX
  - Substitute "dark cycles" for the initial 14 bases
- If dark cycles are not used, trim the first 14 bases from each read before mapping.



E. coli DNA NGS

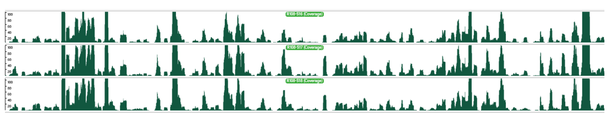
The characteristics of PicoPLEX-012.2 were documented using triplicate fifteen picogram aliquots of DNA from E. coli K-12 MG1655 (Coriell). The libraries were synthesized and amplified using the Rubicon PicoPLEX DNA-seq library prep kit and sequenced on the MiSeq v2. The final results were analyzed in DNAnexus.platform and DNAnexus.classic

NGS statistics (R108)

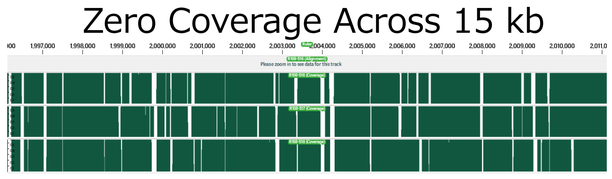
number of reads	mapped reads	read length	average insert	chimeric reads	base mismatch	average GC	mean coverage	bases covered	human reads
1.1 M	96%	80 bp	205 bp	0.7%	0.3%	48.8%	21.2 X	87.4%	<0.001%

PicoPLEX was shown to sequence E. coli very efficiently, with <6% unmapped reads, 0.7% chimeric reads and <1% unassigned bar codes. In addition, even at 21X coverage there is ~90% coverage. The highly-biased coverage is a key property of PicoPLEX that is critical for reproducible human aneuploidy detection with minimal coverage, and should also be very useful in identification of bacterial and viral genomes.

NGS Coverage Across 15 kb (R108)



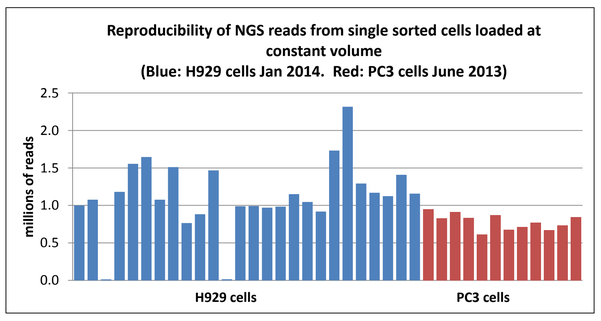
Zero Coverage Across 15 kb



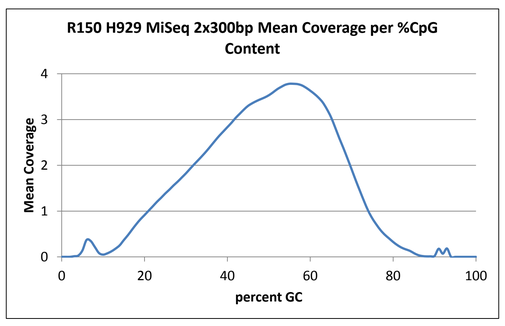
Results: NGS coverage is very reproducible library to library. Deep sequencing up to 100X is possible for a small fraction of the genome, whereas there is still reproducible, albeit very low coverage over most of the genome. The reproducible low coverage is also achieved with single cells, and can be useful for genotyping and mutation detection using PCR as an assay with much higher dynamic range.

Reproducible Read Yield Without Quantification

Equal volumes of twelve single-cell PicoPLEX-012.2 libraries made from flow sorted PC3 cells (bar codes 1-6) and flow-sorted PBMCs (bar codes 7 – 12) were pooled at constant volume, AMPure XP purified, and sequenced on a HiSeq 2000.



Very Good GC Coverage of PicoPLEX



Single Human Cancer Cell NGS

Human PBMC and clonally-expanded PC3 prostate cancer cells were flow sorted into 96-well plates containing 5 uL buffer. A bar-coded PicoPLEX DNA-seq library was synthesized and amplified from 6 single cells of each type, pooled, cleaned up with AMPure XP, diluted and sequenced on a MiSeq v2. After trimming the initial 14 bases of quasi-random primer sequence the reads were mapped using BWA-MEM, processed in Picard_Mark_Duplicates, and further characterized in DNAnexus.classic.

NGS statistics (R57)

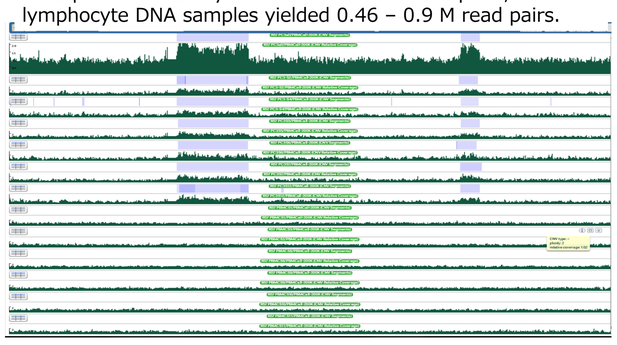
Human cells	Input	Number of reads	Mapped reads	Read length (bp)	Unassigned bar codes	low quality reads
PBMC (R57.S7-12)	Single cells	0.9 – 1.4M	96.9%- 97.4%	84	<1%	0.04 %– 0.06%
PC3 (R57.S1-6)	Single cells	0.52 – 1.5M	97.0 %- 97.4%	84	<1%	0.03% - 0.05%

Human cells	Insert size (bp)	Chimeric reads	Base mismatch rate	Average GC	Mean coverage	bases covered
PBMC (R57.S7-12)	200–255	0.53%– 0.58%	0.25%-0.50%	42.2% –42.6%	0.025X – 0.05X	1.8% – 3.1%
PC3 (R57.S1-6)	191-255	0.49%– 0.52%	0.50%–1.0%	42.3% –42.8%	0.015 X – 0.042X	1.1% – 2.0%

Results: The PBMC and PC3 cell NGS statistics were very similar, as also is the case at low coverage for mouse, E. coli, and other genomes. These results are expected because the quasi-random sequences have not been "humanized" or in any other way dependent upon the presence of specific sequences.

Reproducible Human CNV Calls Using Single PC3 cells (R57)

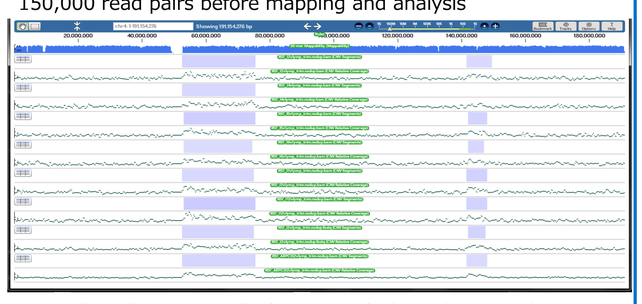
Equal volumes of six single-cell PicoPLEX-012.2 libraries from flow-sorted PC3 cells and six 15 pg lymphocyte DNA samples were made into PicoPLEX-012.2 libraries, amplified, pooled with equal volume, AMPure purified, and sequenced on a MiSeq v2. PC3 cells yielded 0.26 – 0.6 M read pairs; lymphocyte DNA samples yielded 0.46 – 0.9 M read pairs.



Results: DNAnexus CNV analysis of chr 4 detected 23 Mbp and 6 Mbp duplicated regions on 4q when the information from all six PC3 cells was pooled and compared with all six lymphocyte controls (trace 1). All PC3 cells (traces 2-6) show 2 aneuploidies, however none of the lymphocyte samples (traces 7 – 13) showed any CNVs.

Accurate Human CNV Calls at 150,000 Read Pairs (R57)

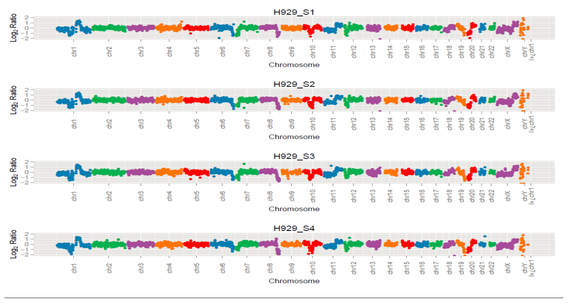
Equal volumes of six single-cell PicoPLEX-012.2 libraries made from flow-sorted PC3 cells and six 15 pg samples of lymphocyte DNA were made into PicoPLEX-012.2 libraries, amplified, pooled with equal volume, AMPure XP purified, and sequenced on a MiSeq v2. The fastq files were randomly downsampled to 150,000 read pairs before mapping and analysis



Results: All six PC3 cells (traces 2-6) show the aneuploidies except for one abnormality that did not meet p-value criteria.

Reproducible Single Flow-Sorted H929 Cell Aneuploidy Determination (250K SE 35 b reads)

Single Flow-Sorted H929 Cell Aneuploidy from 100K SE 35 b reads)



CNV in Day 5 Human Embryos

(Dr. Brian Mariani, Genetics and IVF Institute)

Fully-consented, frozen, day-5 embryos were lysed using PicoPLEX DNA-seq, aliquoted into triplicate 5-cell portions, amplified and sequenced on the MiSeq v2. Up to 3 embryos were processed from each case, and up to 3 libraries from each embryo. Equal volume PicoPLEX library aliquots were labeled and hybridized to BlueGnome 24sure arrays. Libraries were pooled in groups with 48 compatible bar codes, and sequenced in a 2x100 MiSeq v2 PE experiment.

NGS statistics (R107)

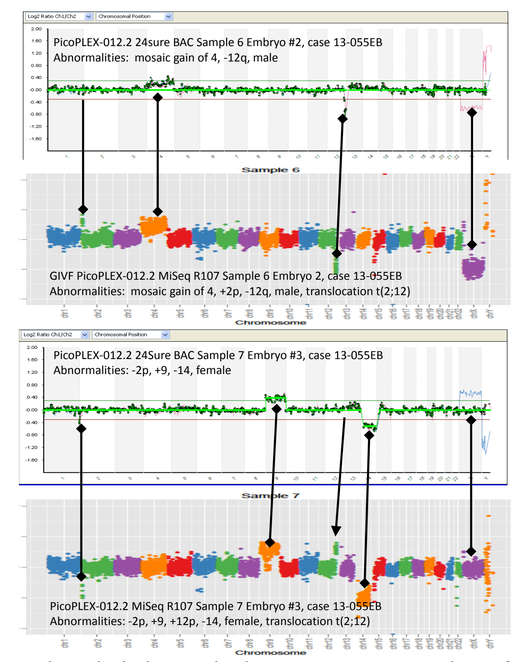
Human samples	DNA input	Number of clusters	Mapped reads	Read length	Unassigned bar codes	low quality reads
embryo aliquots	~ 5-cell	127K – 718K	98% - 99%	97.5 bp	<1%	0.04% – 0.05%

Human samples	Average insert size (bp)	library duplicates (%)	Est. library size	chimeric reads (%)	Base mismatch rate (%)	average GC (%)	Mean coverage (%)	bases covered (%)
embryo aliquots	250 – 270 bp	0.7% - 1.5%	22M 105M	0.36% - 0.48%	0.3% - 0.5%	43.5% - 44.1%	0.08 X	1.1%– 1.5%

Results: Data were mapped with BWA-MEM. PicoPLEX performed at GIVF, University of Leuven,

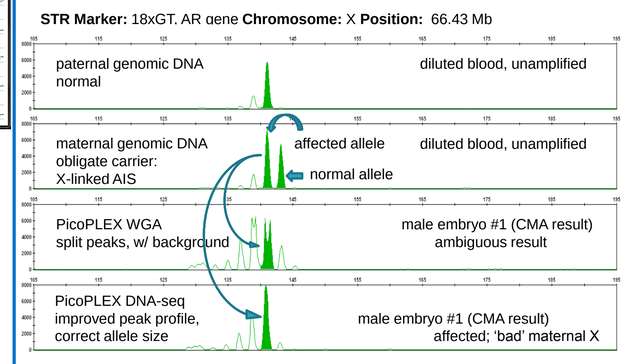
Comparison of Aneuploidy Calls From BAC Arrays to NGS (R107)

BlueGnome 24sure BAC arrays are the gold standard of IVF genetic testing worldwide, using the BlueGnome SurePLEX (PicoPLEX WGA) kits. The PicoPLEX-012.2 kits under development were tested for equivalency to 24sure arrays by using the same libraries for array and NGS applications. The panels below show chromosome copy number analyses from embryo #2 and #3 of case 13-055EB.



Results: Black diamonds show apparent gain or loss of chromosomes or segments of chromosomes. Embryo #2 shows +2p and -12p by NGS but only -2p by BAC analysis. Embryo #3 shows -2p and +12p by NGS but only -2p by BAC analysis. NGS is consistent with a translocation t(2:12). All other results are the same for NGS and BAC analysis. The presence of a maternal reciprocal translocation t(2:12) was confirmed by FISH analysis of the mother's blood. The exchange of 20 Mbp of chr2 and 6 Mbp of chr12 is likely responsible for infertility of the couple, and had not been detected by conventional karyotyping and FISH.

STR Haplotyping to Detect a Single-Gene Disorder, PicoPLEX WGA vs PicoPLEX DNA-seq



Conclusions about PicoPLEX

- PicoPLEX single-cell libraries have excellent sequence quality.
- PicoPLEX DNA-seq NGS libraries have a very simple and fast workflow that is suitable CNV screening, resequencing, and de novo sequencing.
- At 0.002X average coverage CNV of 10-20 Mbp size are reproducibly detected in single cells
- At 0.002X average coverage, PicoPLEX DNA-seq is able to detect aneuploidy in IVF materials with comparable sensitivity to the 24sure BAC arrays with less than 200K reads allowing more than 100 embryos to be screened in less than one day using a MiSeq.
- At 0.002X coverage identification of signature sequences should be useful for metagenomics, forensics, and infectious disease testing.
- At 20X average coverage 90% genome coverage is possible, enabling construction of haplotypes and scaffolds of large complex genomes.
- Whole exome and other targeted sequencing should enable the genotype, SNV, and structural variants to be determined from PicoPLEX DNA-seq libraries.