Single-cell genome sequencing

Mario Nuvolone

Technical Journal Club

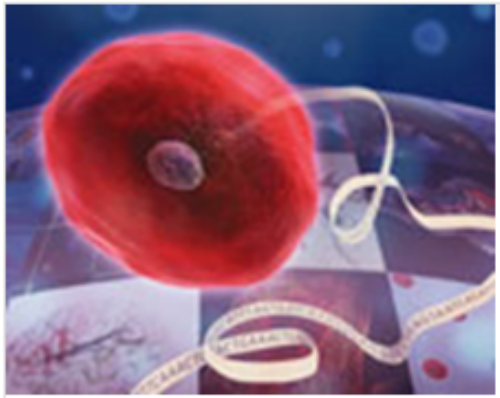
28th January 2013

University Hospital Zurich
Institute of Neuropathology

nature methods
Techniques for life scientists and chemists

SPECIAL FEATURE

Method of the Year 2013



Special Feature

Nature Methods' choice for Method of the Year 2013 is single-cell sequencing. A collection of articles present the unique considerations related to sequencing single cells and highlight recent applications in biology and medicine. The Methods to Watch feature provides a look at possible future Methods of the Year.

EDITORIAL

Method of the Year 2013

Methods to sequence the DNA and RNA of single cells are poised to transform many areas of biology and medicine.

METHOD OF THE YEAR | COMMENTARY | SPECIAL FEATURE

Dissecting genomic diversity, one cell at a time

Paul C Blainey & Stephen R Quake

Emerging technologies are bringing single-cell genome sequencing into the mainstream; this field has already yielded insights into the genetic architecture and variability between cells that highlight the dynamic nature of the genome.

METHOD OF THE YEAR | NEWS FEATURE | SPECIAL FEATURE

Singled out for sequencing

Single-cell genome and transcriptome sequencing methods are generating a fresh wave of biological insights into development, cancer and neuroscience. Kelly Rae Chi reports.

SPECIAL FEATURE | COMMENTARY | METHOD OF THE YEAR

Entering the era of single-cell transcriptomics in biology and medicine

Rickard Sandberg

Recent technical advances have enabled RNA sequencing (RNA-seq) in single cells. Exploratory studies have already led to insights into the dynamics of differentiation, cellular responses to stimulation and the stochastic nature of transcription. We are entering an era of single-cell transcriptomics that holds promise to substantially impact biology and medicine.

METHOD OF THE YEAR | COMMENTARY | SPECIAL FEATURE

The promise of single-cell sequencing

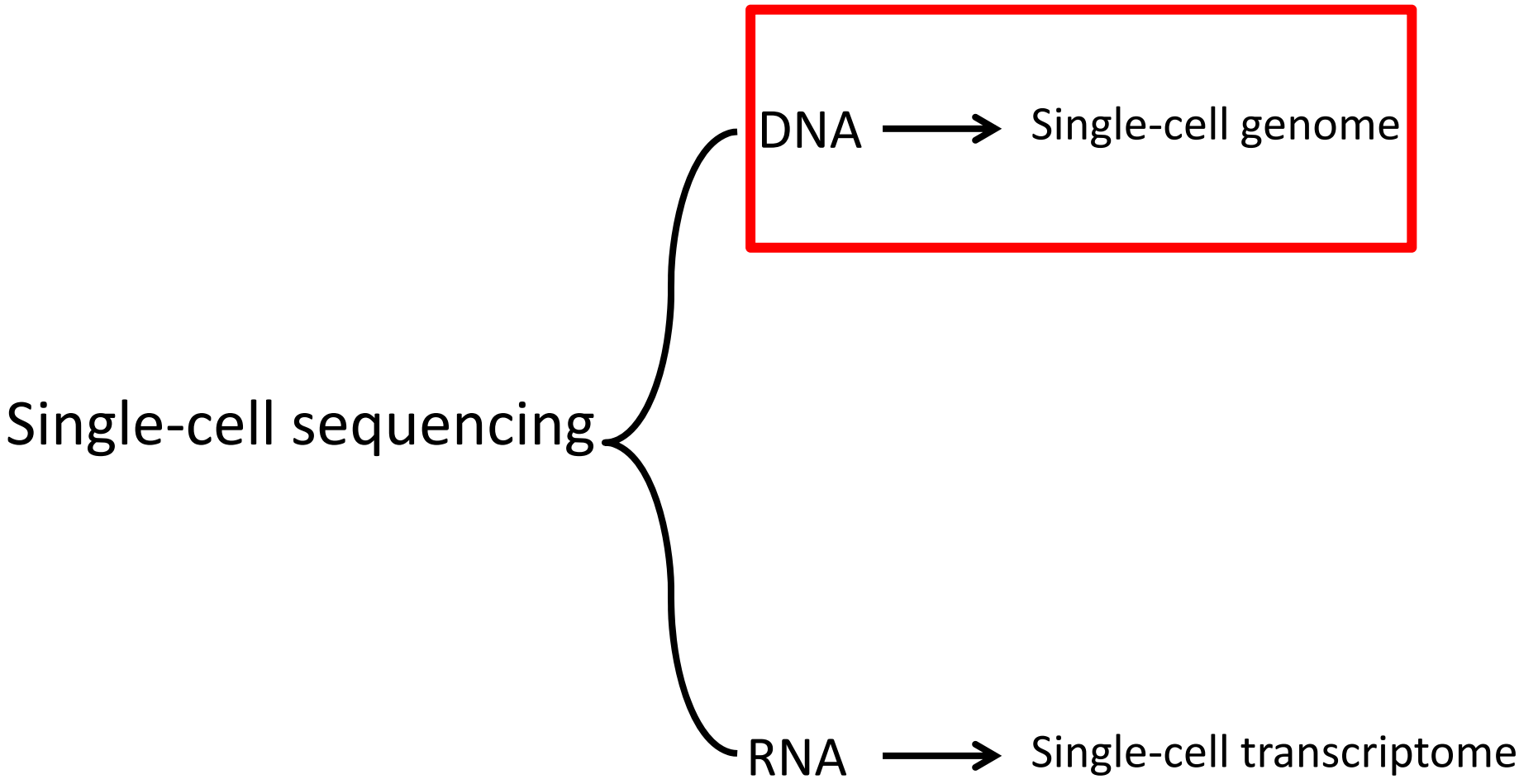
James Eberwine[1,2], Jai-Yoon Sul[1], Tamas Bartfai[3] & Junhyong Kim[2,4]

Individual cells of the same phenotype are commonly viewed as identical functional units of a tissue or organ. However, the deep sequencing of DNA and RNA from single cells suggests a more complex ecology of heterogeneous cell states that together produce emergent system-level function. Continuing development of high-content, real-time, multimodal single-cell measurement technologies will lead to the ultimate goal of understanding the function of an individual cell in the context of its microenvironment.

SPECIAL FEATURE | PRIMER

Single-cell sequencing

A brief overview of how to derive a genome or transcriptome from a single cell.


Single-cell sequencing
DNA
Single-cell genome
RNA
Single-cell transcriptome

Single-cell genome sequencing: Why/When?

Current applications:

1. Microorganisms that cannot be cultured

2. Genome heterogenity in tumors

3. Genome heterogenity among different cells of multicellular organisms

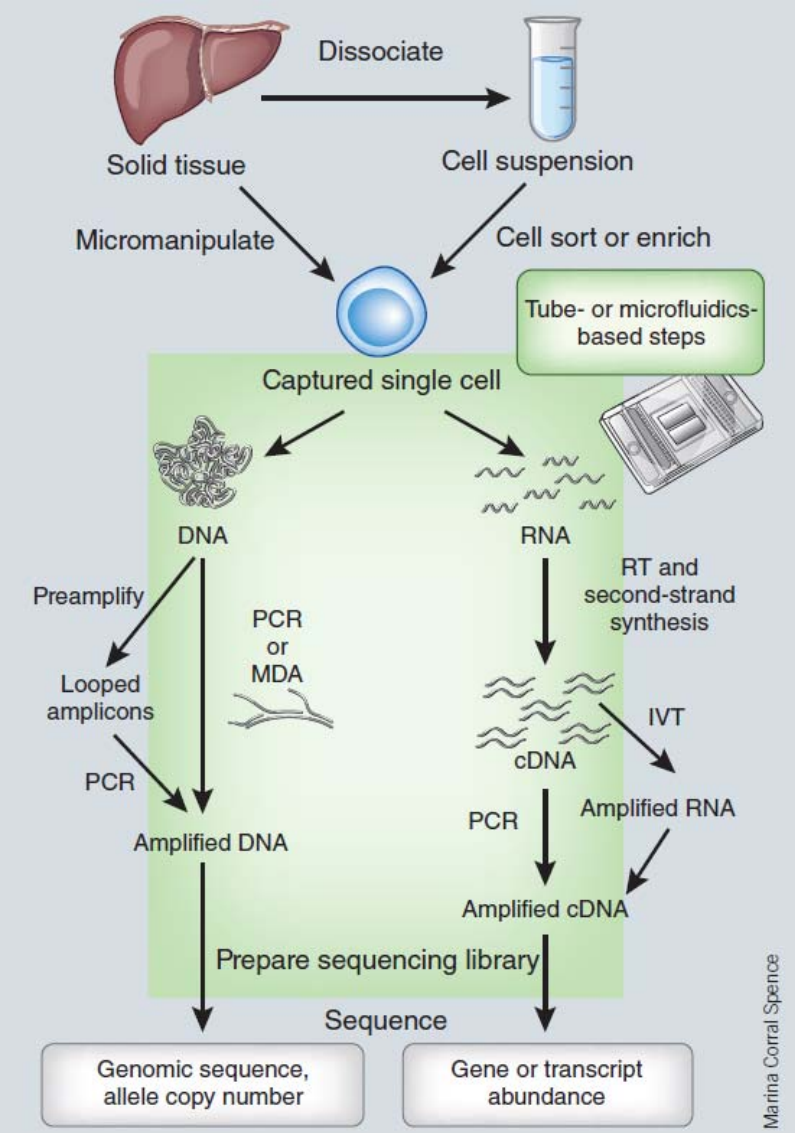
4. Genome heterogenity among different cells of a cell line

5. Gametogenesis and prenatal genetic diagnosis/screening

Single-cell genome sequencing: How?

Main steps:

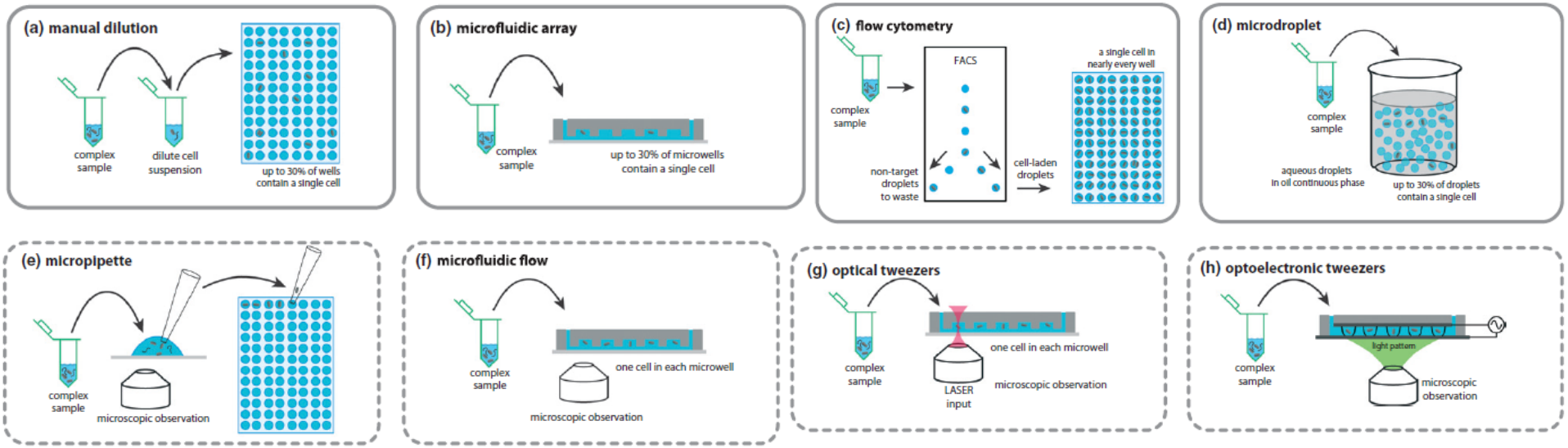
1. Single cell isolation and lysis
2. DNA amplification
3. DNA sequencing
4. Data analysis



Workflows for amplifying and sequencing the RNA or genomic DNA of a cell. MDA, multiple displacement amplification; RT, reverse transcription; IVT, *in vitro* transcription.

Nawy. Nat Methods Special feature: Method of the Year 2013

Single cell isolation

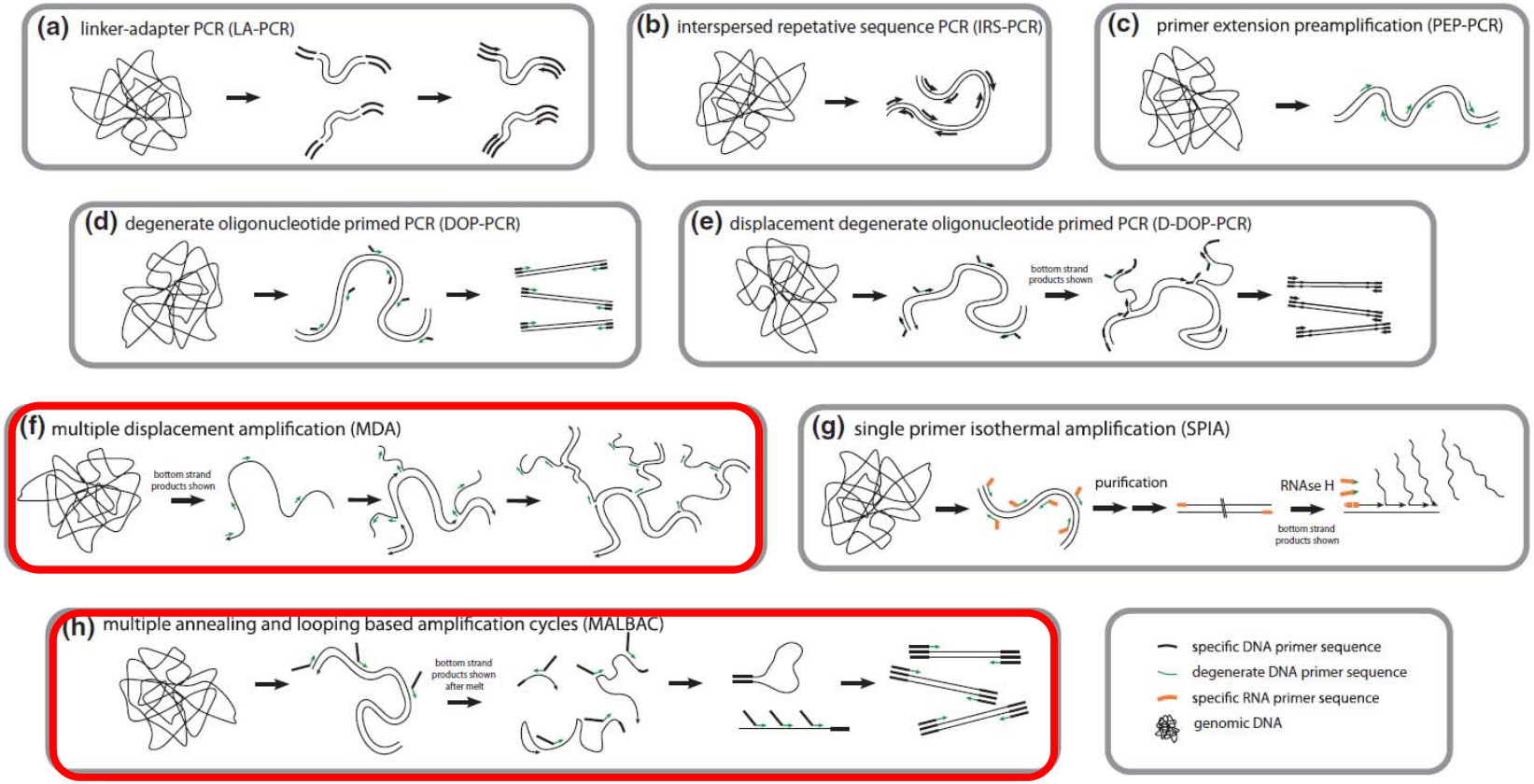


(i) Cell isolation approaches

	implementation	technology platform	commercially available	cell characterization	sampling volume	isolation statistic	cell throughput	potential for parallelization	process volume (reagent cost)	cell/product recovery	notes
random encapsulation	**manual dilution**	standard pipetting	yes	typically none	microliter	poisson	tens	high	microliter (high)	microwell plate	no specialized instrumentation required
	microfluidic array	microfabricated chambers	yes	fluorescence & micrograph possible	nanoliter	poisson	hundreds	low	nanoliter (low)	custom procedure	straightforward, rapid; recovery of products is challenging
	flow cytometry (FACS)	droplets in air	yes	optical scattering, fluorescence intensity	nanoliter	poisson (picked)	hundreds	low	microliter (high)	microwell plate	rapid, cells experience high shear, requires many cells
	microdroplet	droplets in oil, passive microfluidics	emerging	fluorescence & micrograph possible	picoliter	poisson	thousands+	high	picoliter (low)	yet to be shown	highest throughput; rxn setup & recovery methods needed
micromaniupulation	**micropipette**	ultra-precise pipetting	yes	fluorescence & micrograph possible	nanoliter	directed	tens	low	microliter (high)	custom procedure	requires open platform
	microfluidic flow	active microfluidics	yes	fluorescence & micrograph possible	picoliter	directed	tens	low	nanoliter (low)	custom procedure or microwell plate	requires precise flow control
	laser tweezers	laser microscopy	yes	fluorescence & micrograph possible	femptoliter	directed	tens	moderate	nanoliter (low)	custom procedure	microfluidic integration has been demonstrated
	optoelectronic tweezers	opto-electrokinetics	emerging	fluorescence & micrograph possible	femptoliter	directed	hundreds - thousands	high	--	yet to be shown	high-throughput potential; fluidic integration needed

Blainey FEMS Microbiol Rev 2013

DNA amplification

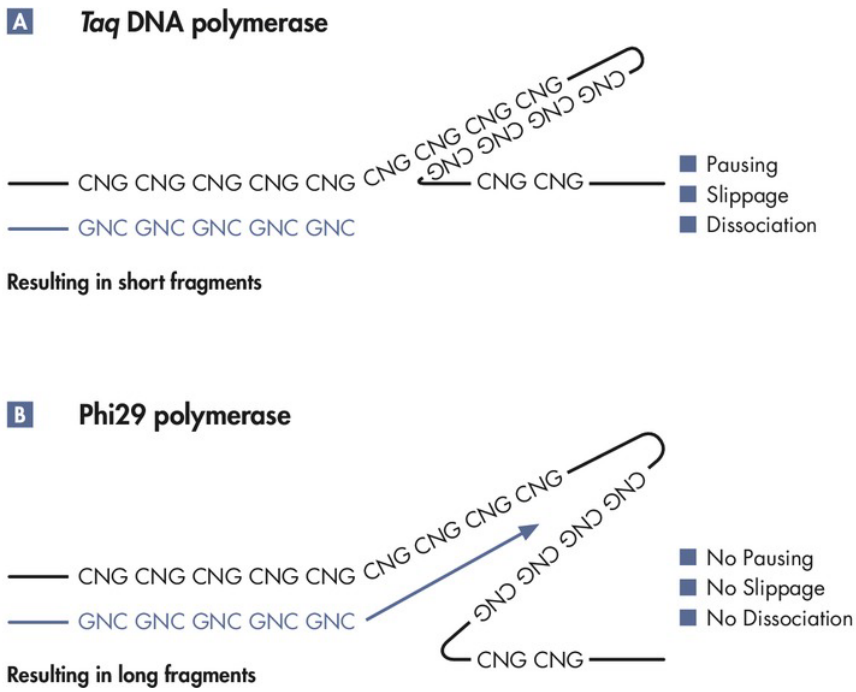


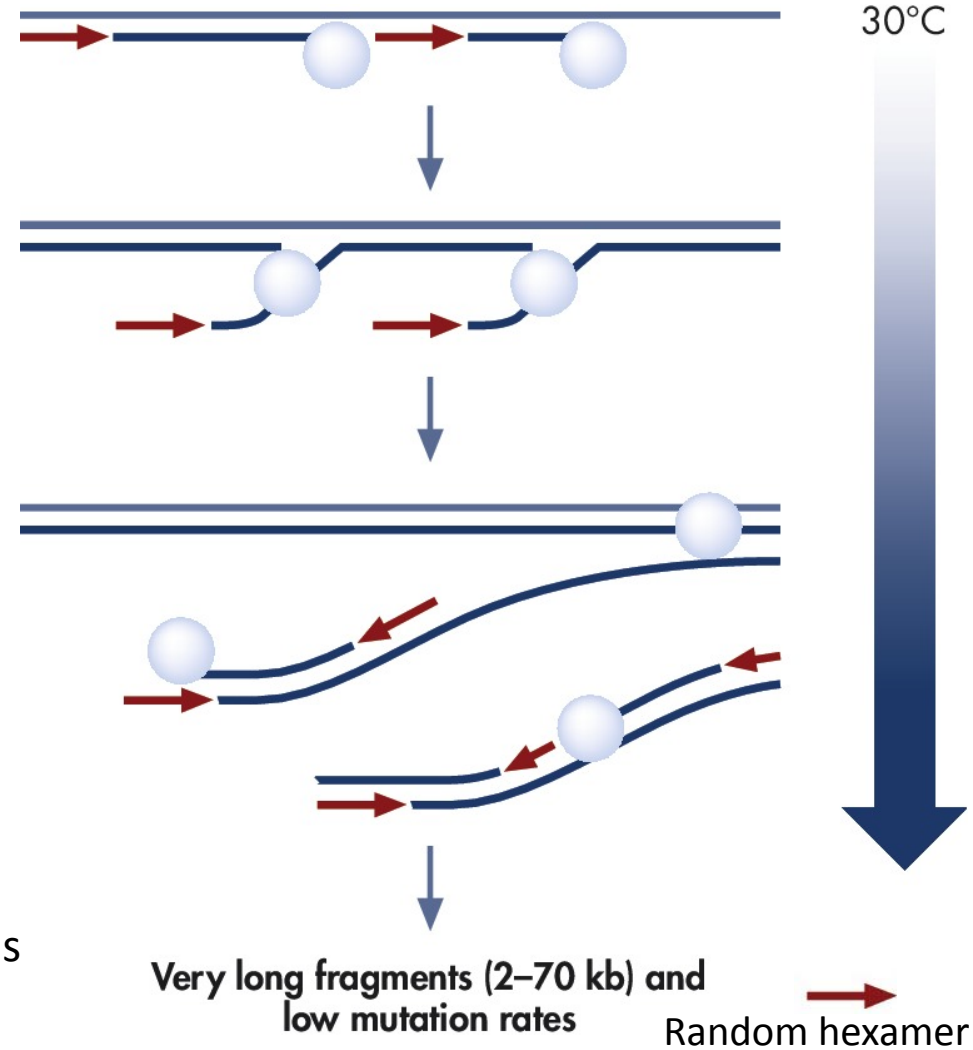
(i) Whole-genome amplification (WGA) methods

WGA method	primer ligation	fully degenerate primers	specific primers	partially degenerate primers	RNA/DNA hybrid primers	strand displacement synthesis	PCR	exponential amplification steps	linear amplification steps	amplification of unknown sequence	bias suppression	commercially marketed	commonly applied for single cells	product length (nt)	product concentration (ng/rxn µL)
LA-PCR	yes	no	yes	no	no	no	yes	1	0	yes	poor	no	no	100–1000	50–100
PEP-PCR	no	yes	no	no	no	no	yes	1	0	yes	poor	no	no	100–1000	1
IRS-PCR	no	no	yes	no	no	no	yes	1	0	no	poor	no	no	100–1000	50–100
DOP-PCR	no	no	no	yes	no	no	yes	1 - 2	0	yes	improved	Sigma	yes	100–1000+	50–100
D-DOP-PCR	no	no	yes	yes	no	yes	yes	2	0	yes	improved	Rubicon	yes	100–1000	50–100
MDA	no	yes	no	no	no	yes	no	1	0	yes	improved	GE, Qiagen	yes	> 10,000	1000
SPIA	no	no	no	yes	yes	no	no	0	1	yes	no data	NuGen	no	100–2000	100–150
MALBAC	no	no	yes	yes	no	yes	yes	1	1	yes	improved	no	emerging	500–1500	100

Blainey FEMS Microbiol Rev 2013

Multiple Displacement Amplification (MDA)





Phi29 polymerase:

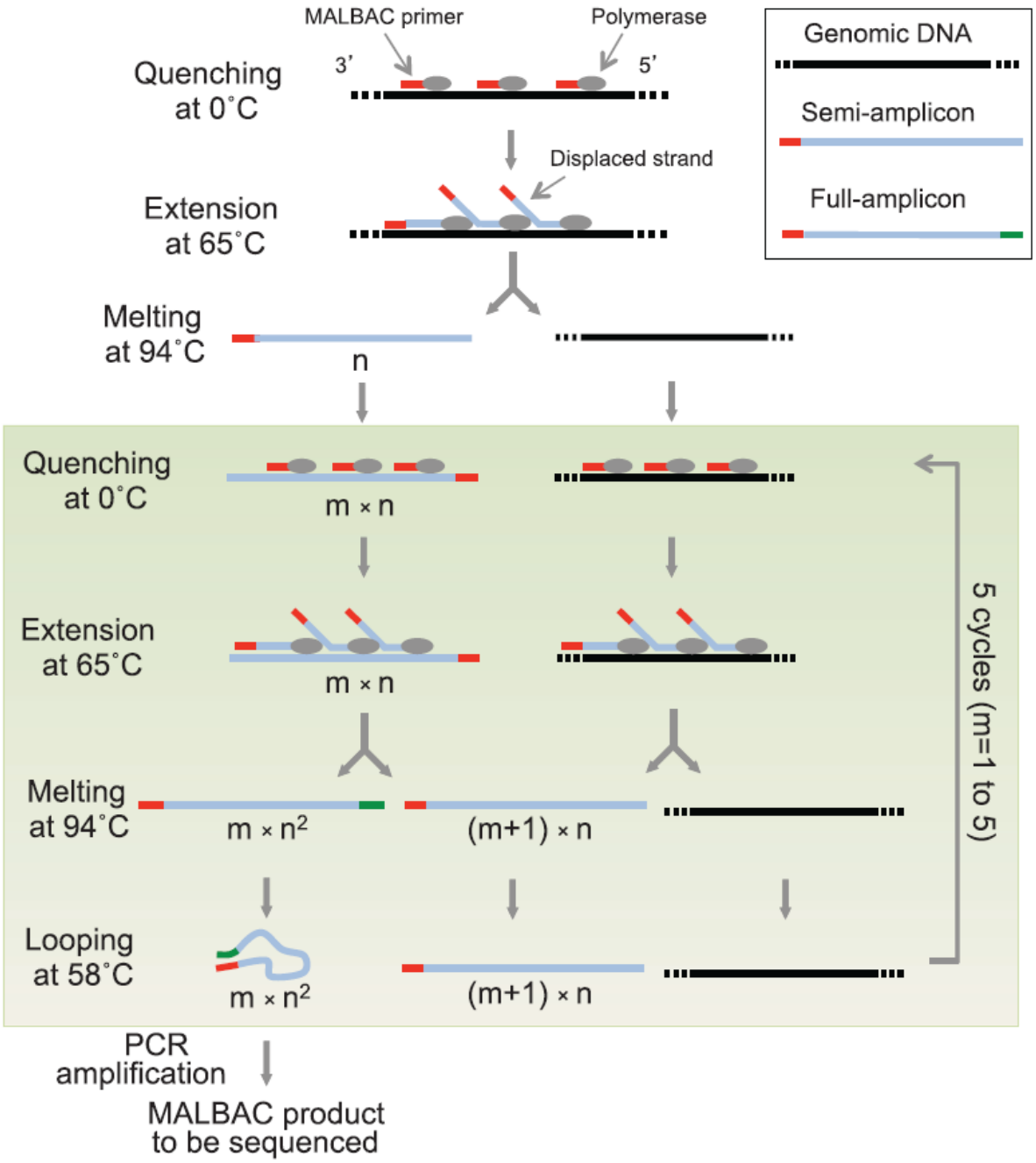
- Creates and displaces overlapping synthesis products from the template as single-stranded DNA
- Works under isothermal conditions
- 3'–5' exonuclease-mediated proofreading
- Strong processivity

Qiagen, Repli-g Single cell kit

Blainey FEMS Microbiol Rev 2013

Dean et al. Proc Natl Acad SCI U S A 2002

Multiple Annealing and Looping-based Amplification Cycles (MALBAC)



MALBAC primers:
Common 27-nt + 8 variable nt
Can evenly hybridize to gDNA at 0°C

DNA polymerase with displacement activity

Five cylces of quasi-linear amplification

Final exponential ampification via PCR using primers annealing to the common sequence of MALBAC primers

Zong et al. Science 2012

LETTER

NATURE | VOL 472 | 7 APRIL 2011

doi:10.1038/nature09807

Tumour evolution inferred by single-cell sequencing

Nicholas Navin[1,2], Jude Kendall[1], Jennifer Troge[1], Peter Andrews[1], Linda Rodgers[1], Jeanne McIndoo[1], Kerry Cook[1], Asya Stepansky[1], Dan Levy[1], Diane Esposito[1], Lakshmi Muthuswamy[3], Alex Krasnitz[1], W. Richard McCombie[1], James Hicks[1] & Michael Wigler[1]

Cell

Cell *155*, 1492–1506, December 19, 2013

Genome Analyses of Single Human Oocytes

Yu Hou,[1,6] Wei Fan,[1,4,6] Liying Yan,[1,6] Rong Li,[1] Ying Lian,[1] Jin Huang,[1] Jinsen Li,[1] Liya Xu,[1] Fuchou Tang,[1,5,*] X. Sunney Xie,[1,2,*] and Jie Qiao[1,3,*]

nature biotechnology

VOLUME 31 NUMBER 12 DECEMBER 2013 NATURE BIOTECHNOLOGY

Massively parallel polymerase cloning and genome sequencing of single cells using nanoliter microwells

Jeff Gole[1], Athurva Gore[1], Andrew Richards[1], Yu-Jui Chiu[2], Ho-Lim Fung[1], Diane Bushman[3], Hsin-I Chiang[1,5], Jerold Chun[3], Yu-Hwa Lo[4] & Kun Zhang[1]

LETTER

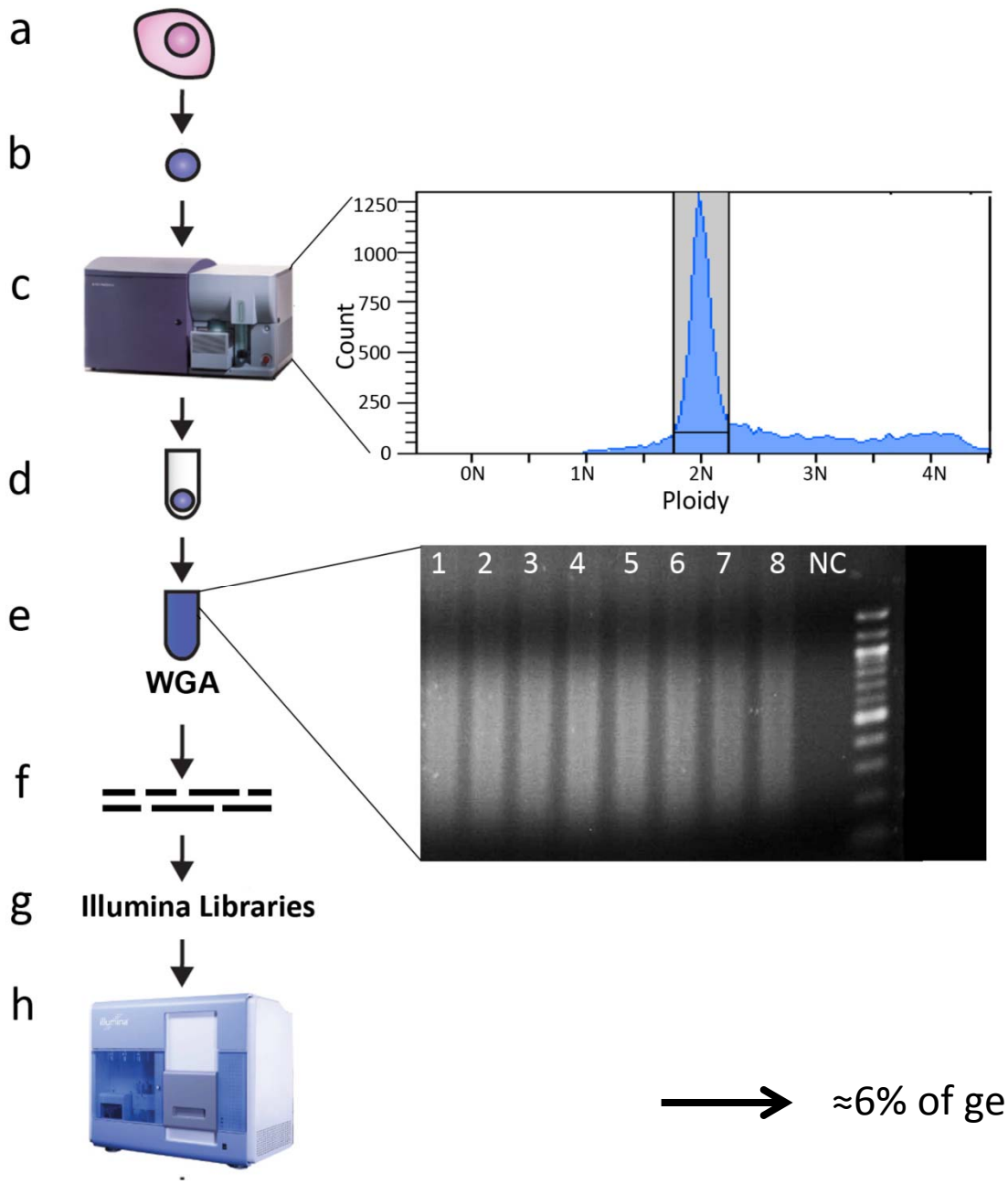
NATURE | VOL 472 | 7 APRIL 2011

doi:10.1038/nature09807

Tumour evolution inferred by single-cell sequencing

Nicholas Navin[1,2], Jude Kendall[1], Jennifer Troge[1], Peter Andrews[1], Linda Rodgers[1], Jeanne McIndoo[1], Kerry Cook[1], Asya Stepansky[1], Dan Levy[1], Diane Esposito[1], Lakshmi Muthuswamy[3], Alex Krasnitz[1], W. Richard McCombie[1], James Hicks[1] & Michael Wigler[1]

Tumor-derived single cell genome sequencing



a) Single cell isolation from tumor

b-d) DAPI-stained nuclei (subpopulations based on ploidy) FACS sorted in 96-well plate with lysis buffer

e) Whole genome amplification

f) Sonication to remove specific 28 bp adapters

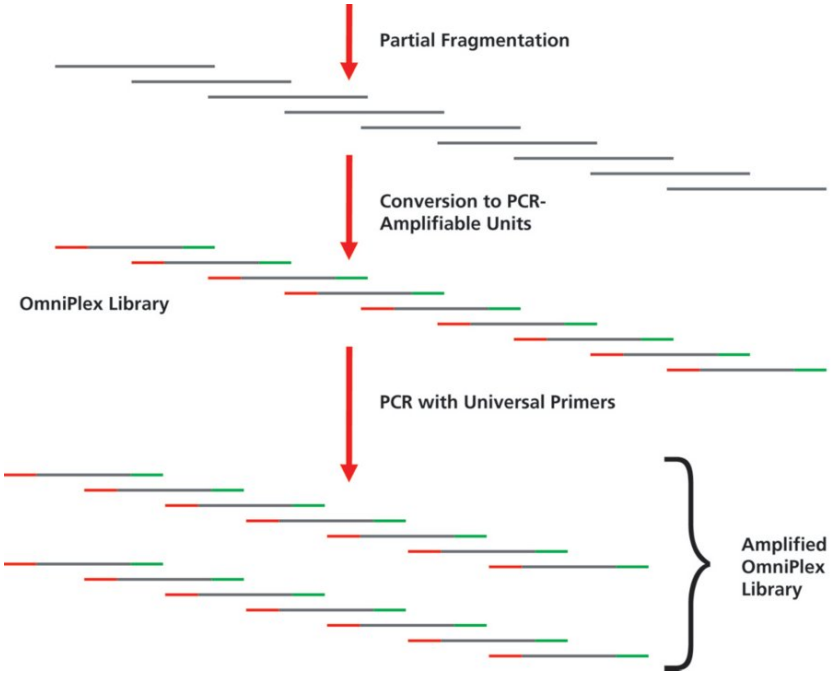
g) Illumina single-end libraries are prepared

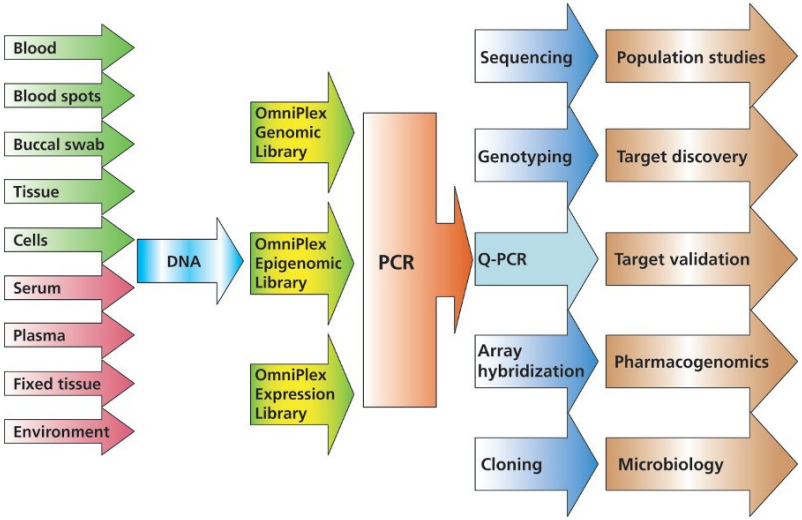
h) Each library is sequenced on individual flow-cell lanes

→ ≈6% of genome coverage per cell

Tumor-derived single cell genome sequencing: WGA

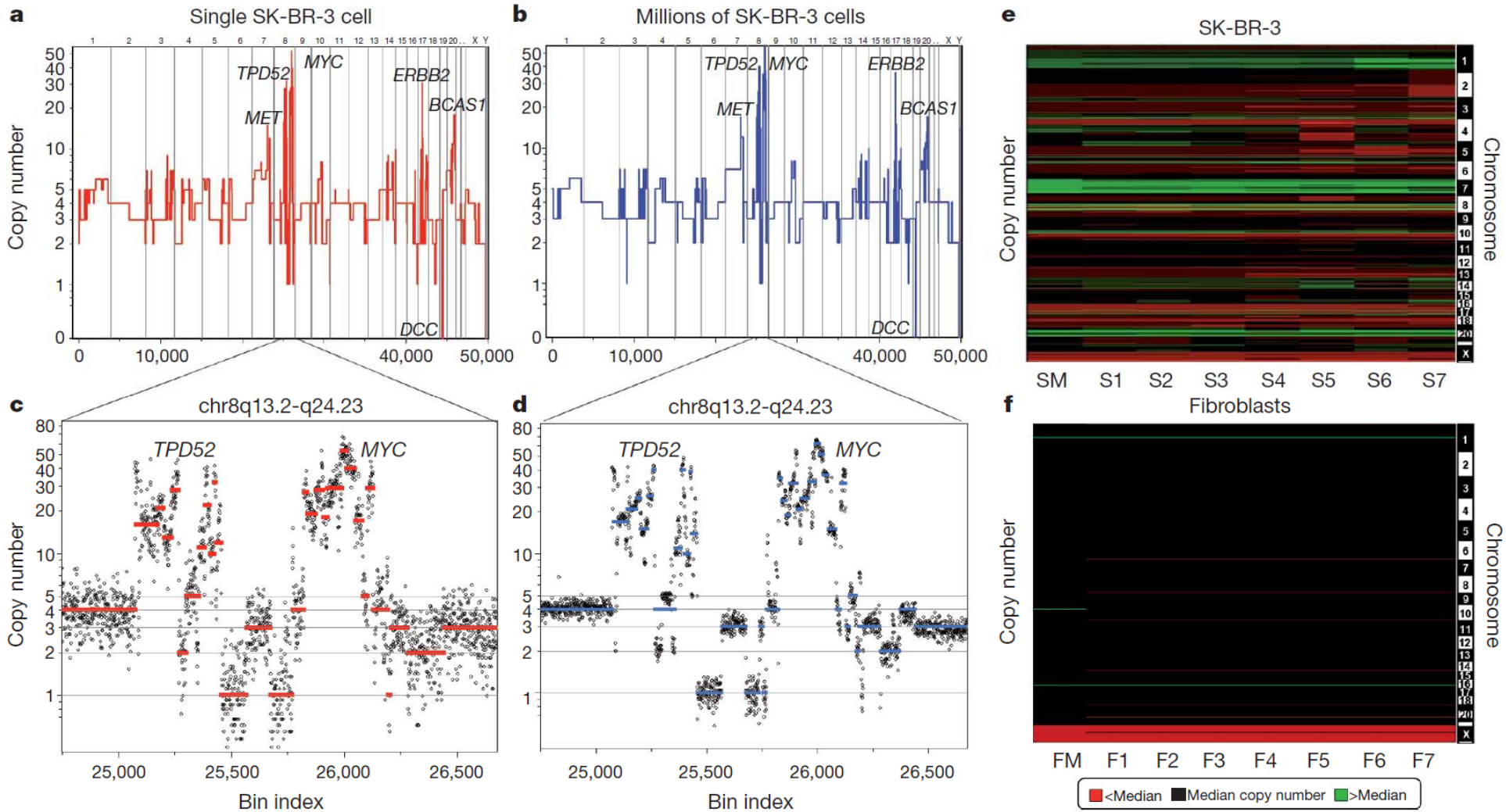
- Through Sigma GenomePlex® WGA kit
- Random fragmentation of gDNA and generation of PCR-amplifiable OmniPlex® Library molecules flanked by universal priming sites
- The OmniPlex® library is then amplified using universal oligonucleotide primers and a limited number of cycles
- Require ng of gDNA with yields of 5-10 µg following PCR
- Suitable with a variety of purified DNA samples (whole blood, buccal swabs, plant sources, FFPE)
- Suitable for various applications (TaqMan®, STR, SNP, Sequencing, CGH, Microarrays)



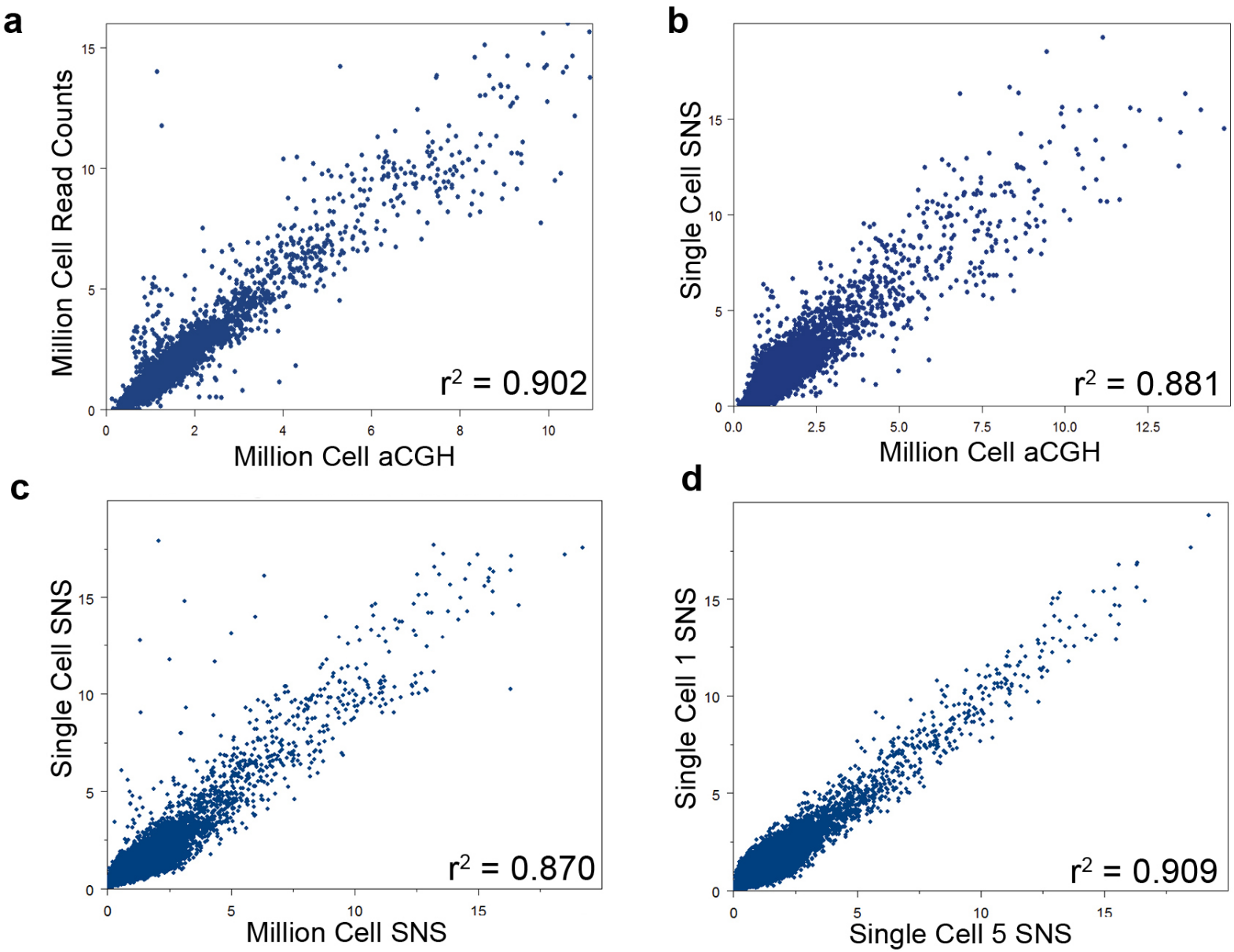


Sigma

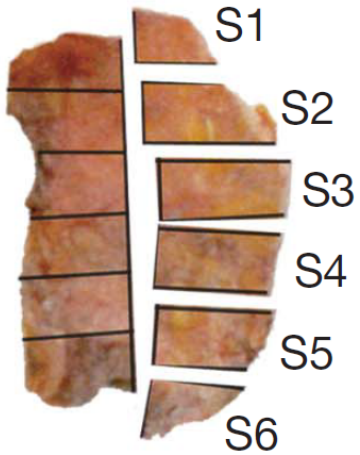
Tumor-derived single cell genome sequencing: validation (I)

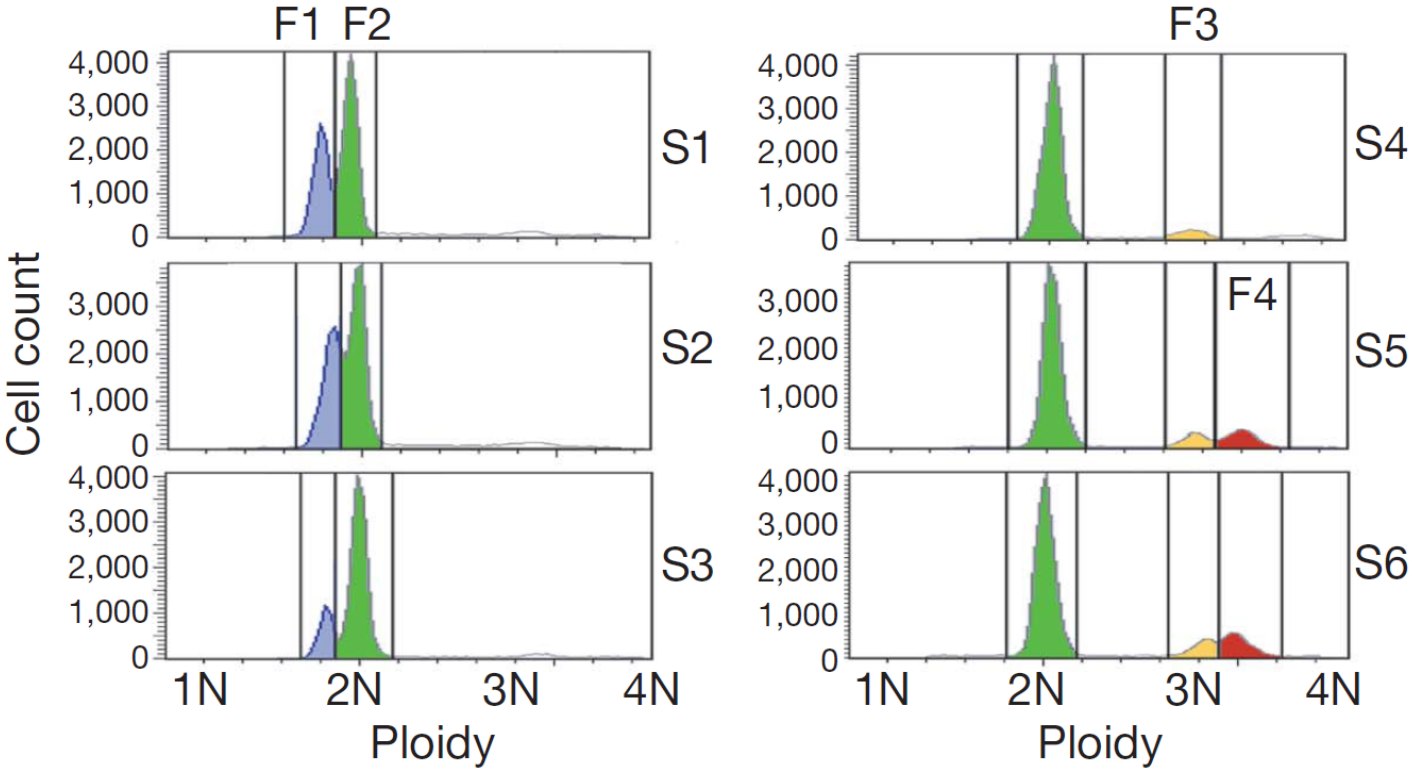


Tumor-derived single cell genome sequencing: validation (II)



Analysis of 100 single cells from a polygenomic breast tumor





Histologic analysis:

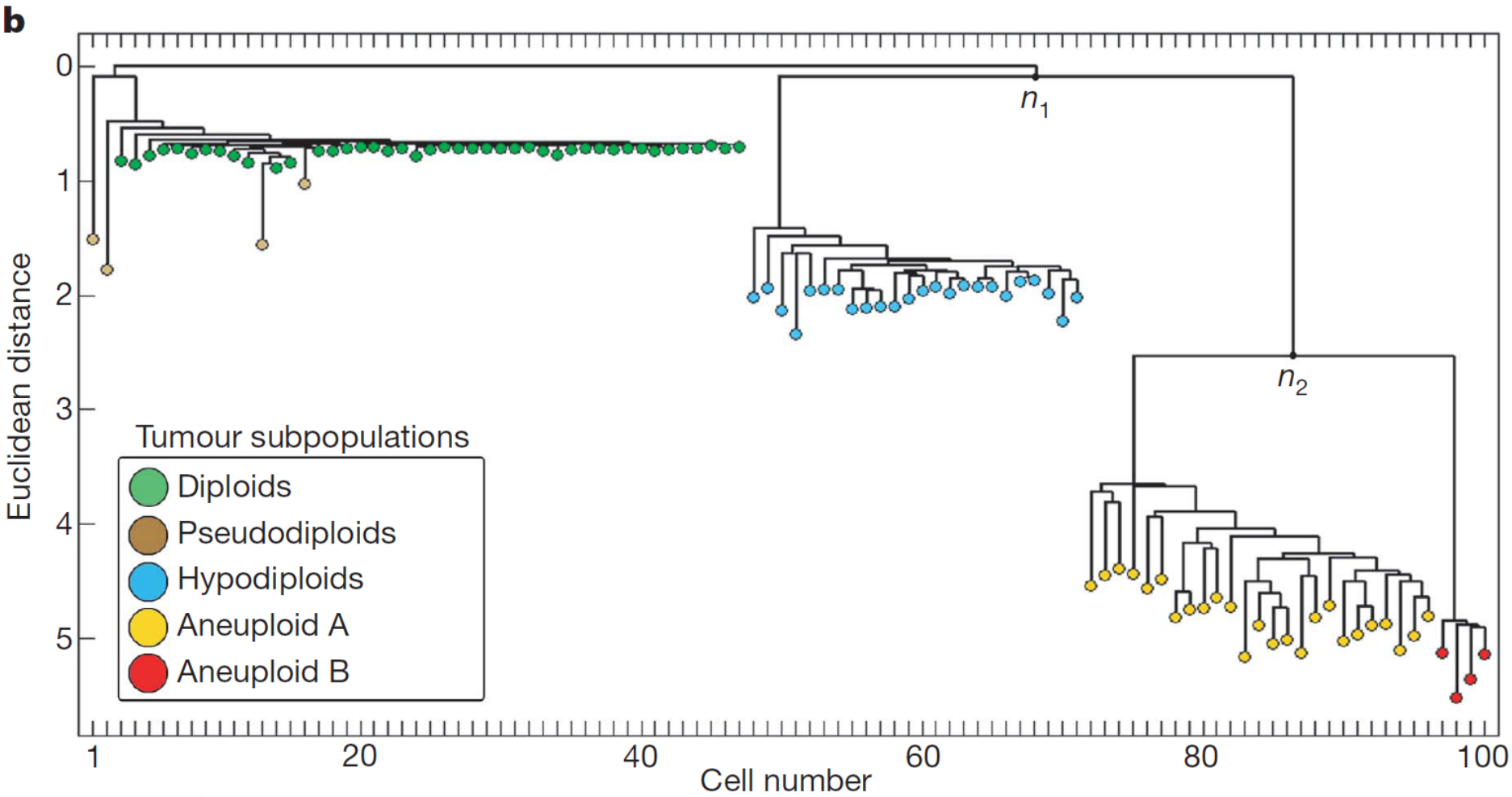
- 63% normal and 37% tumour cells
- Heavily infiltrated with leukocytes

FACS analysis:

- Hypodiploid fraction (F1) exclusive to sectors 1–3
- Diploid 2N fraction (F2) in all sectors (mainly lymphocytes)
- Two sub-tetraploid fractions (F3 andF4) in sectors 4–6

→ 100 single cells from multiple sectors and ploidy fractions sequenced.

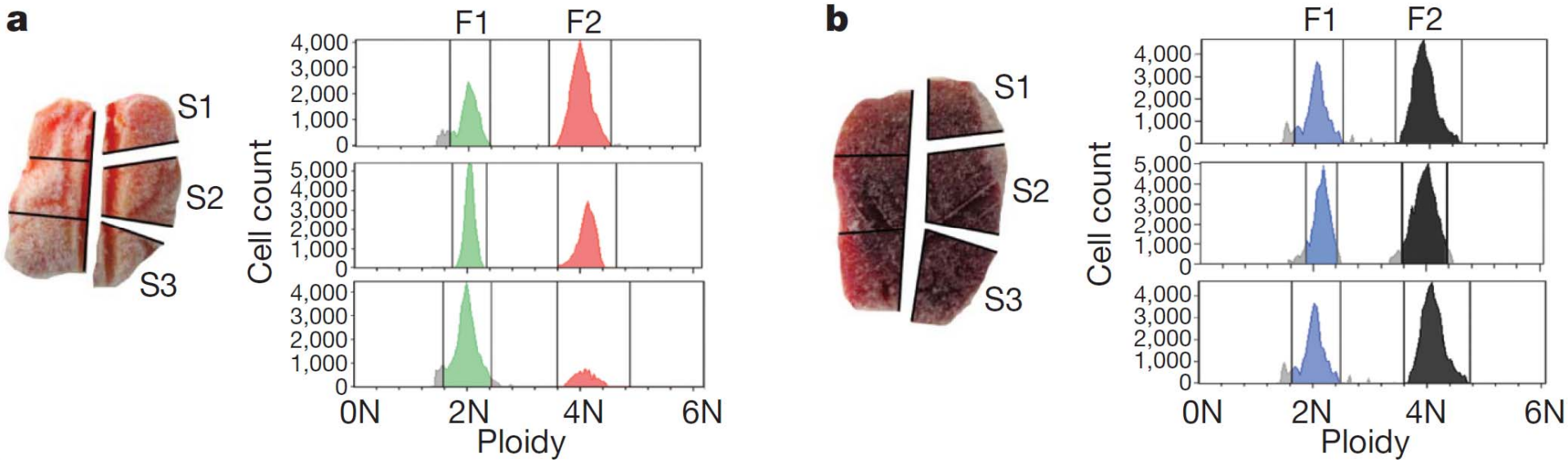
Analysis of 100 single cells from a polygenomic breast tumor



- 3 major 'advanced' tumour subpopulations (H, AA and AB) with highly clonal with complex genomic rearrangements, probably representing three clonal expansions

- Each subpopulation is related to the others by many shared genomic alterations, but they have also diverged and developed distinct attributes

Analysis of 100 single cells from a monogenomic breast tumor and its liver metastasis



Histologic analysis:

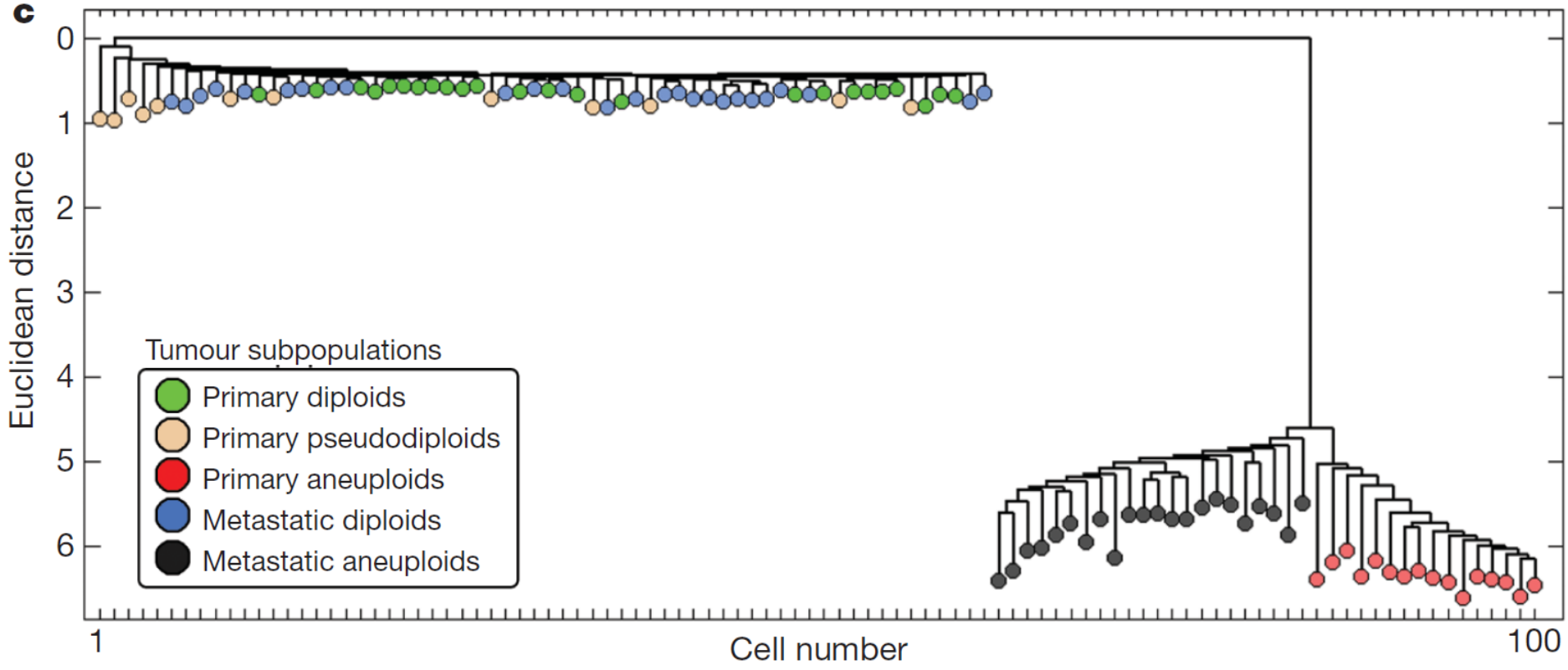
- 50% normal and 50% tumour cells
- Low leukocyte infiltration

FACS analysis:

- Diploid 2N fraction (F1) both in primary and metastatic tumor in all sectors
- Tetraploid fraction (F2) both in primary and metastatic tumor in all sectors

→ 100 single cells from multiple sectors and ploidy fractions sequenced.

Analysis of 100 single cells from a monogenomic breast tumor and its liver metastasis



- Primary tumour formed by a single clonal expansion of an aneuploid cell

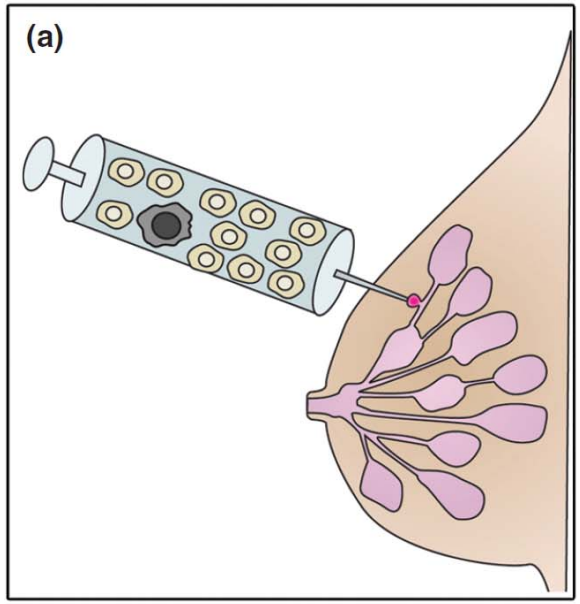
- One of the cells from this primary tumor subsequently seeded the metastatic tumour with little further evolution

Tumor single cell sequencing – summary and future applications

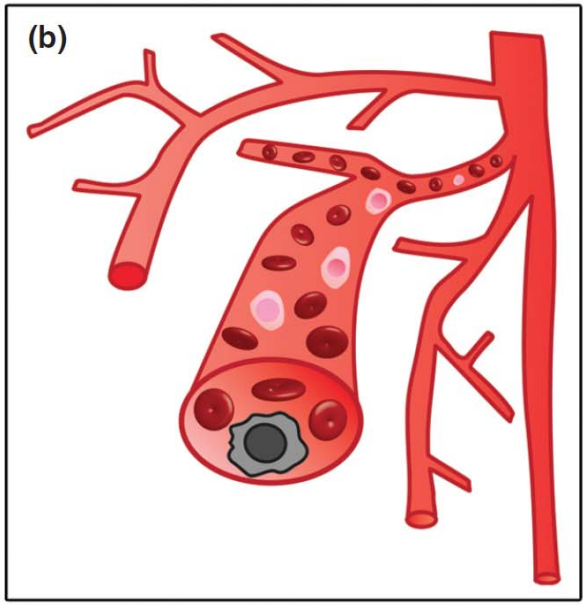
Current achievements:

- Proof-of-principle for the study of tumor genomic heterogeneity and possible evolution

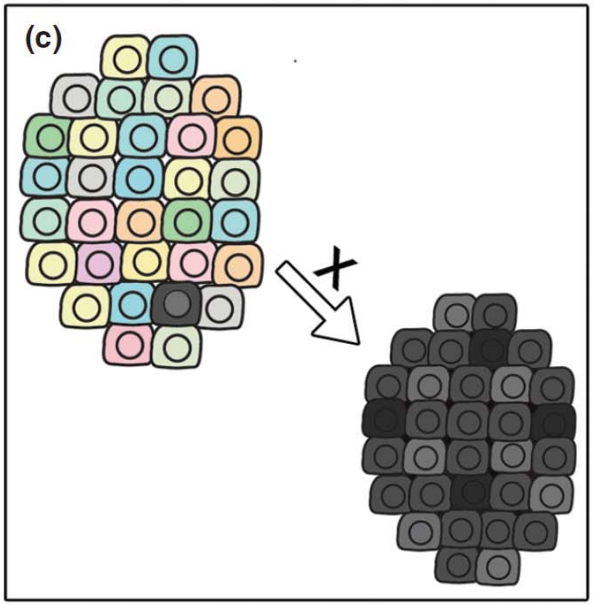
Future applications:



Scarce clinical samples



Circulating tumour cells



Rare chemo resistant cells

Navis & Hicks. Genome Med 2011

PROTOCOL

| VOL.7 NO.6 | 2012 | NATURE PROTOCOLS

Genome-wide copy number analysis of single cells

Timour Baslan[1,2], Jude Kendall[1], Linda Rodgers[1], Hilary Cox[1], Mike Riggs[1], Asya Stepansky[1], Jennifer Troge[1], Kandasamy Ravi[1], Diane Esposito[1], B Lakshmi[3], Michael Wigler[1], Nicholas Navin[4,5] & James Hicks[1]

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA. [2]Molecular and Cellular Biology Program, Stony Brook University, Stony Brook, New York, USA. [3]Ontario Institute for Cancer Research, Toronto, Ontario, Canada. [4]Department of Genetics, University of Texas MD Anderson Cancer Center, Houston, Texas, USA. [5]Department of Bioinformatics and Computational Biology, University of Texas MD Anderson Cancer Center, Houston, Texas, USA. Correspondence should be addressed to J.H. (hicks@cshl.edu).

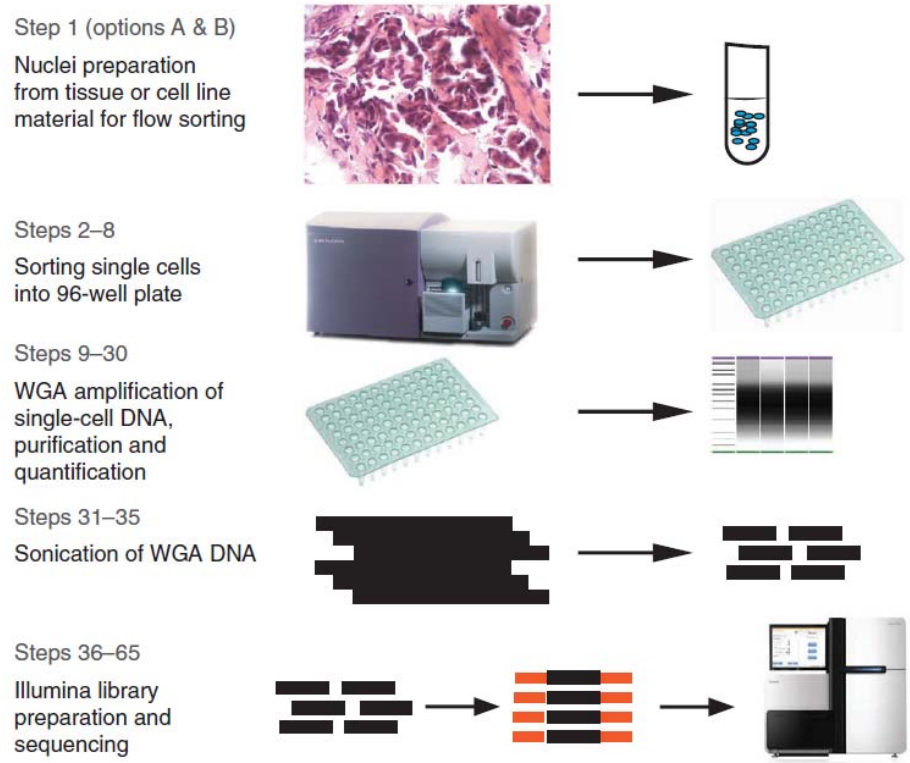


Figure 1 | Schematic of the experimental workflow of SNS. Step numbering corresponds to the Steps of the PROCEDURE. The FACSAria image is courtesy of Becton, Dickinson and Company; reprinted with permission. HiSeq2000 image is courtesy of Illumina.

Cell

Cell *155*, 1492–1506, December 19, 2013

Genome Analyses of Single Human Oocytes

Yu Hou,[1,6] Wei Fan,[1,4,6] Liying Yan,[1,6] Rong Li,[1] Ying Lian,[1] Jin Huang,[1] Jinsen Li,[1] Liya Xu,[1] Fuchou Tang,[1,5,*] X. Sunney Xie,[1,2,*] and Jie Qiao[1,3,*]

[1]Biodynamic Optical Imaging Center, College of Life Sciences and Center for Reproductive Medicine, Third Hospital, Peking University, Beijing 100871, China

[2]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA

[3]Key Laboratory of Assisted Reproduction, Ministry of Education and Beijing Key Laboratory of Reproductive Endocrinology and Assisted Reproductive Technology, Beijing 100191, China

[4]Peking-Tsinghua Center for Life Science, Beijing 100084, China

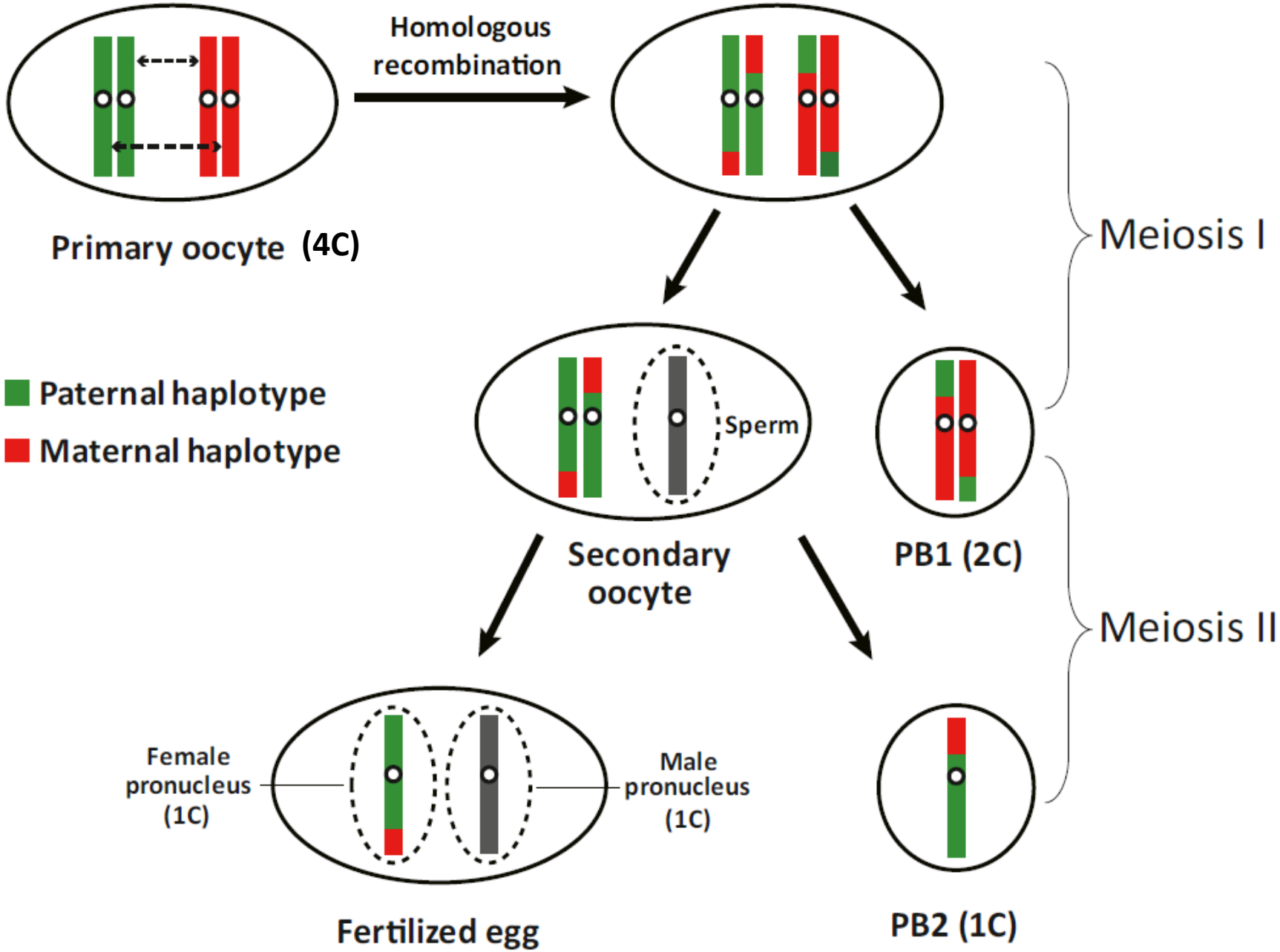
[5]Ministry of Education Key Laboratory of Cell Proliferation and Differentiation, Beijing 100871, China

[6]These authors contributed equally to this work

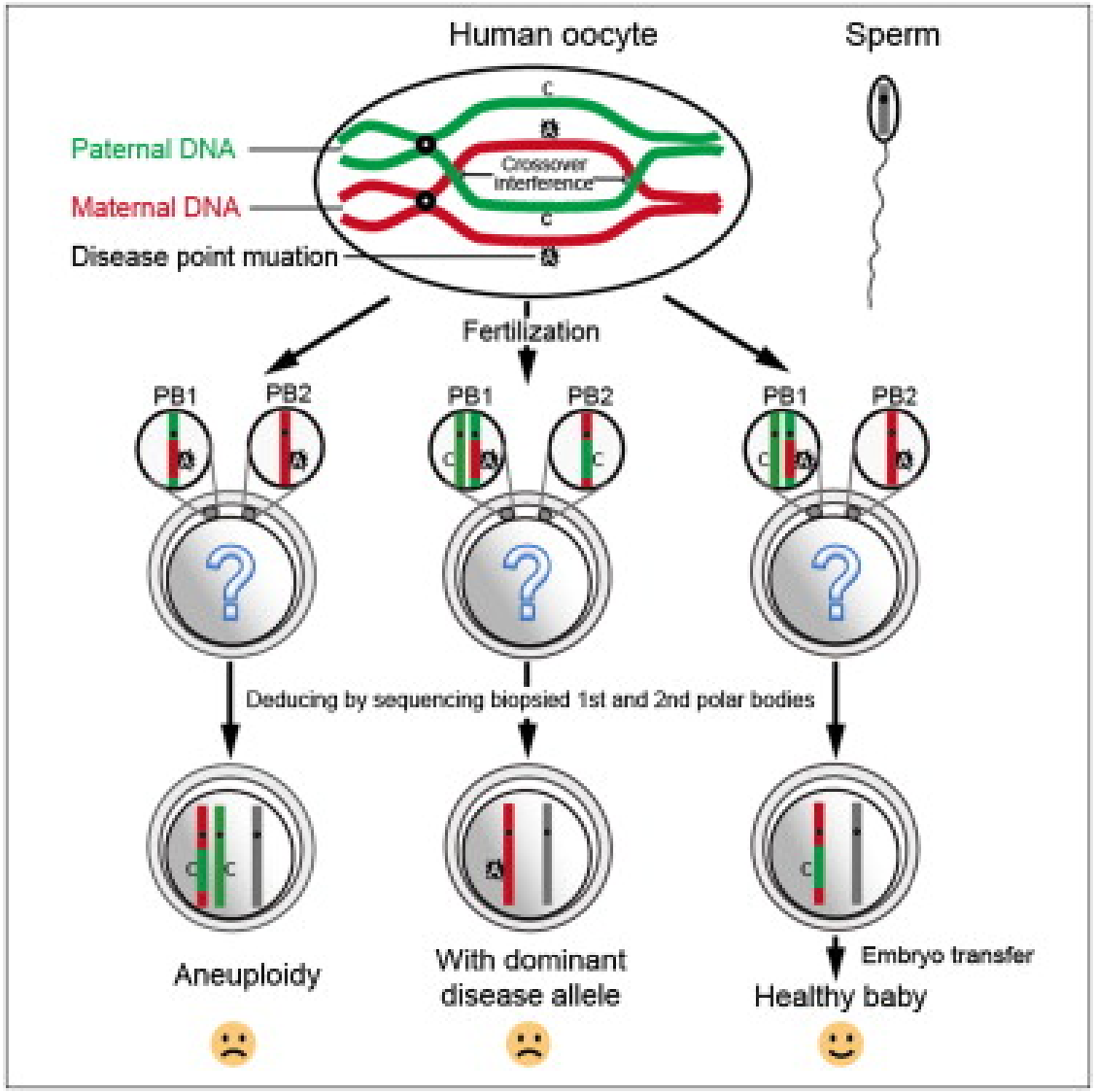
*Correspondence: tangfuchou@pku.edu.cn (F.T.), xie@chemistry.harvard.edu (X.S.X.), jie.qiao@263.net (J.Q.)

http://dx.doi.org/10.1016/j.cell.2013.11.040

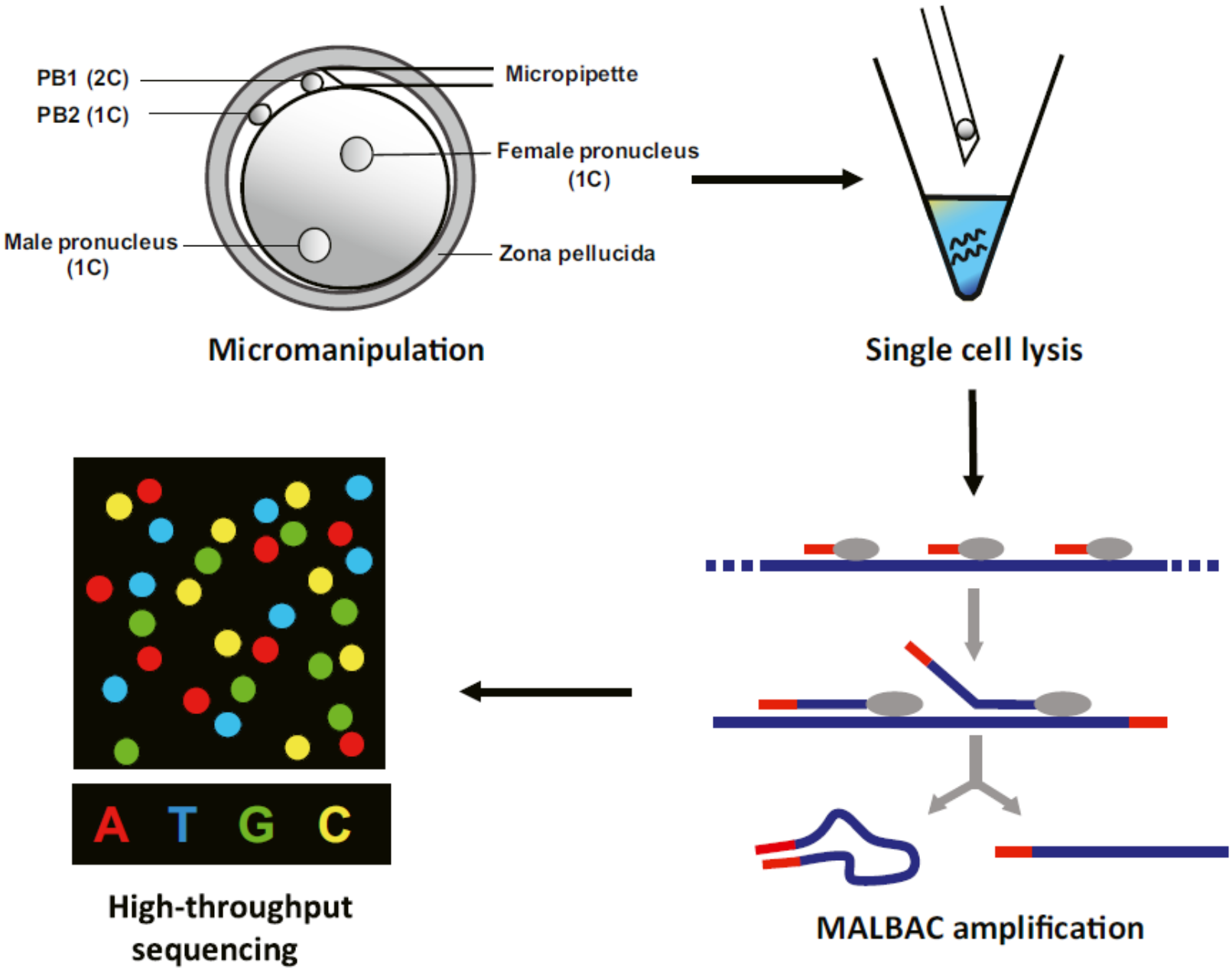
Human oocyte meiosis



MALBAC of single human oocytes: goal



Single-oocyte MALBAC sequencing

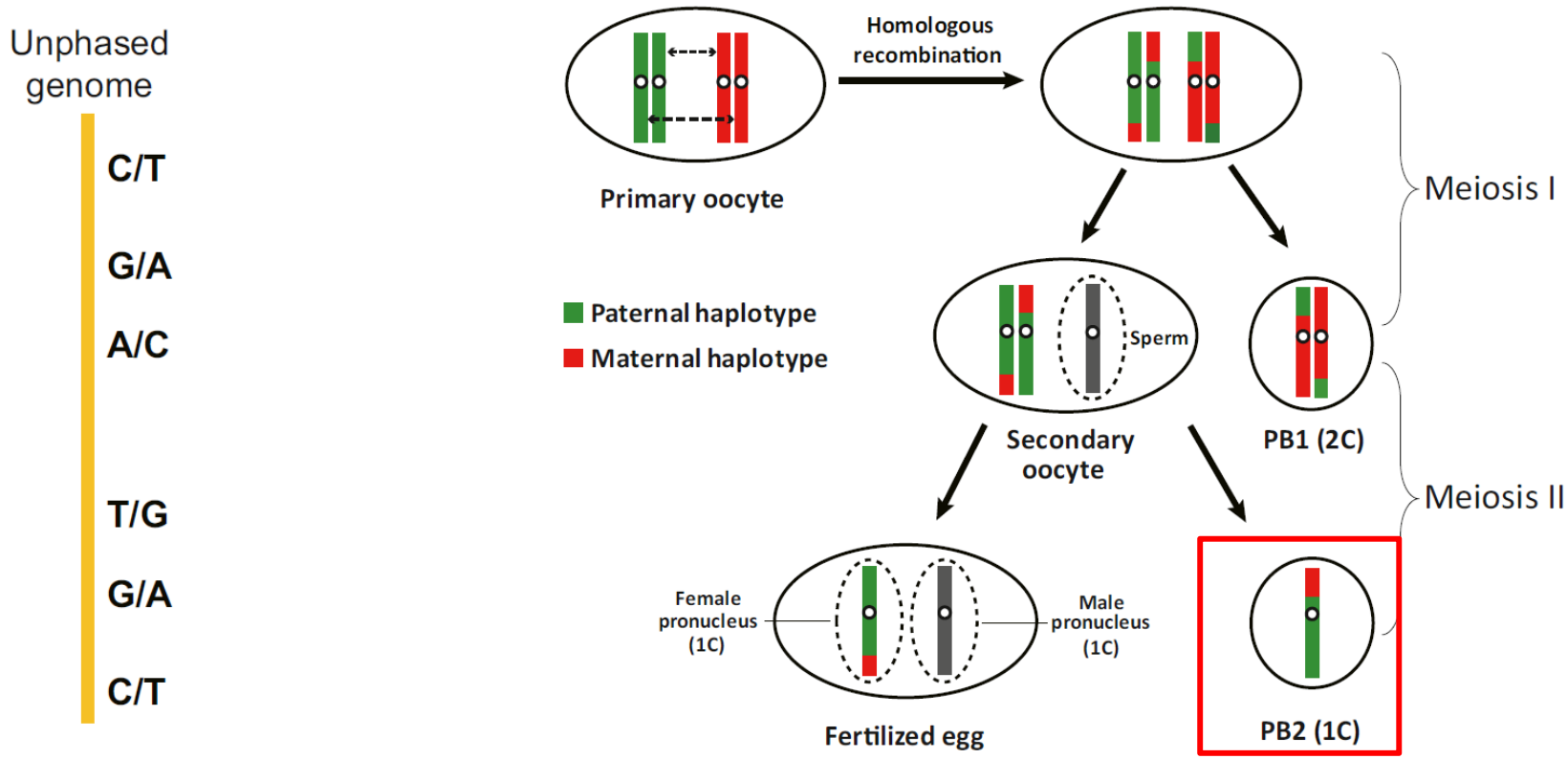


Summary of sequencing information

Donor ID	No. of fertilized oocytes	No. of sequenced polar bodies and pronuclei	Average sequencing depth for each single cell (X)	Average genome coverage	Summed sequencing depth for each donor (X)	Summed genome coverage	No. of phased hetSNPs for each donor	No. of phased SNPs for each single cell
S01	14	35	0.9	35.6%	31.6	93.8%	1,092,055	559,567
S02	5	14	0.4	22.5%	5.4	77.4%	512,620	247,807
S03	6	15	0.4	22.1%	5.8	79.3%	512,149	241,603
S04	4	11	0.4	20.5%	3.9	69.6%	307,583	168,007
S05	12	31	0.9	34.2%	26.5	94.2%	1,046,939	527,501
S06	8	20	0.7	30.8%	14.1	90.1%	721,263	428,383
S07	6	16	0.8	32.8%	12.1	88.4%	675,195	452,225
S08	15	41	0.9	36.5%	36.7	96.6%	1,284,846	621,709
All / Avg	70	183	0.7	31.7%	17.0	86.2%	769,081	405,850

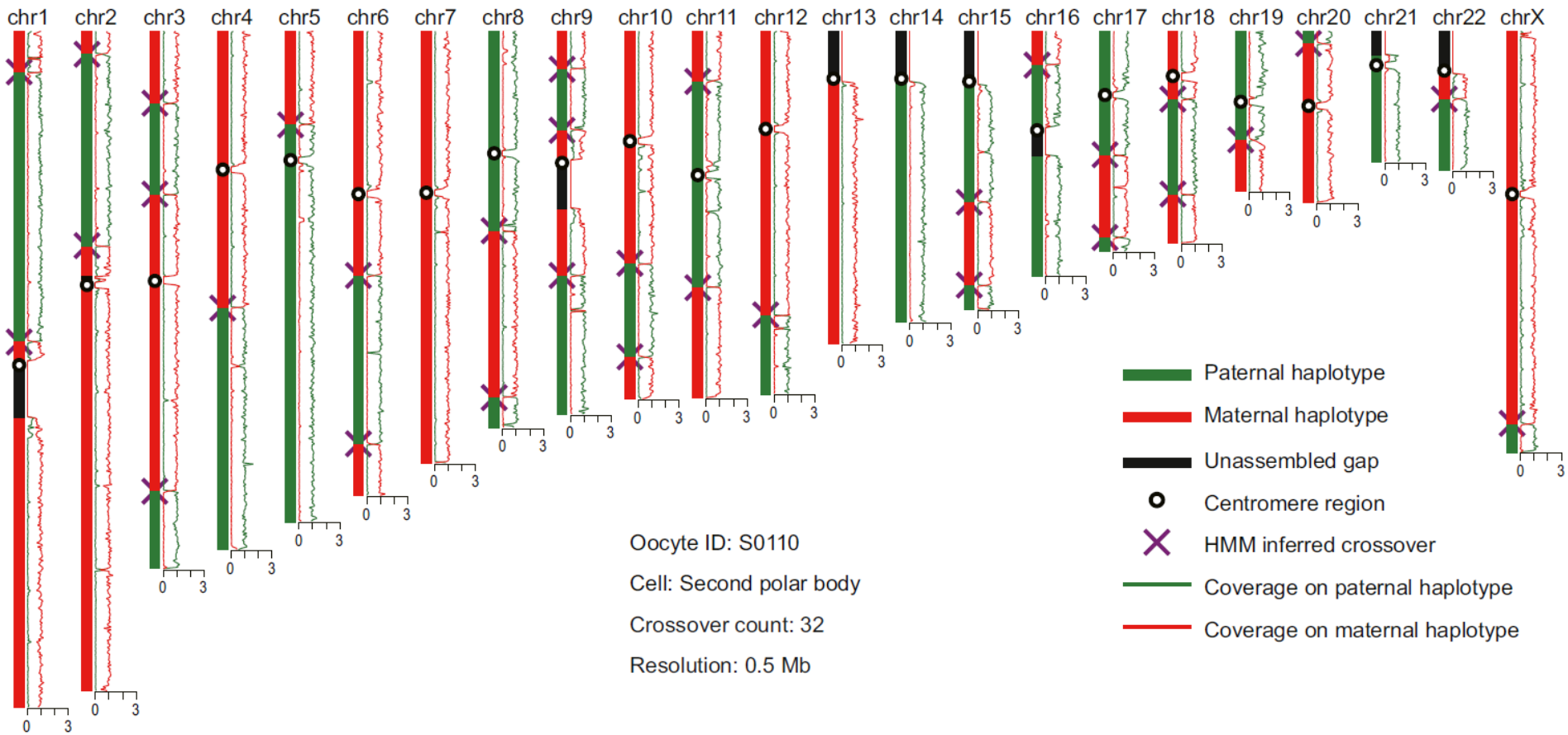
- Oocytes from 8 young healthy donors
- Fertilization performed by intracytoplasmic sperm injection (ICSI)
- First and second polar bodies (PB1 and PB2) and, for verification, female pronuclei (FPN) isolated with laser-assisted micromanipulation
- hetSNPs determined using single-cell sequencing data of all oocytes of a donor

Phasing of donor's hetSNPs



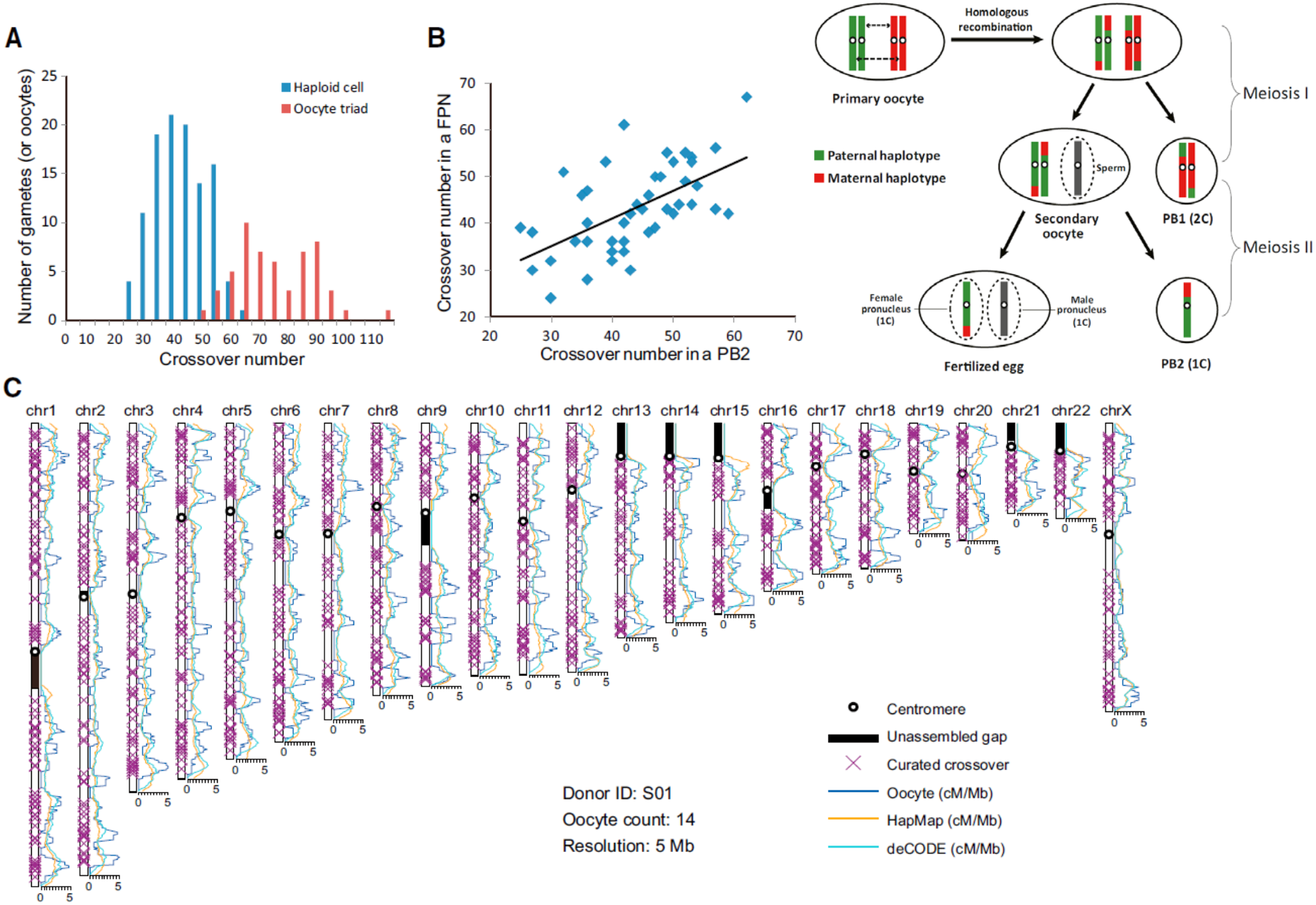
- Haplotype determined based onhaploid PB2 using two independent algorythms
- 91-95% hetSNPs confidentially phased for each donor at the chromosome level

Crossover inference

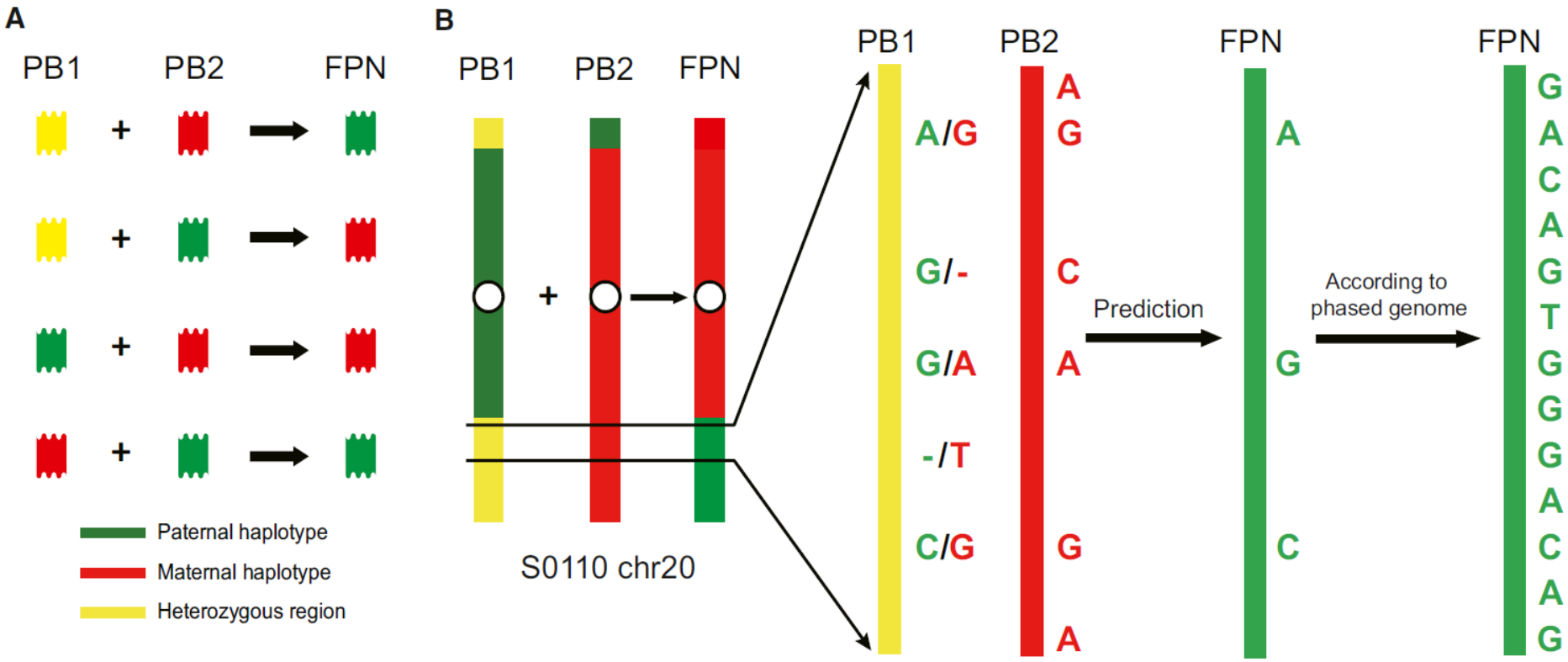


- Crossovers on each chromosome were inferred based on the phased haplotype and on sequencing data using a hidden Markov Model

Distribution characteristics of crossovers in human oocytes



Deduction of female pronucleus haplotype: principle

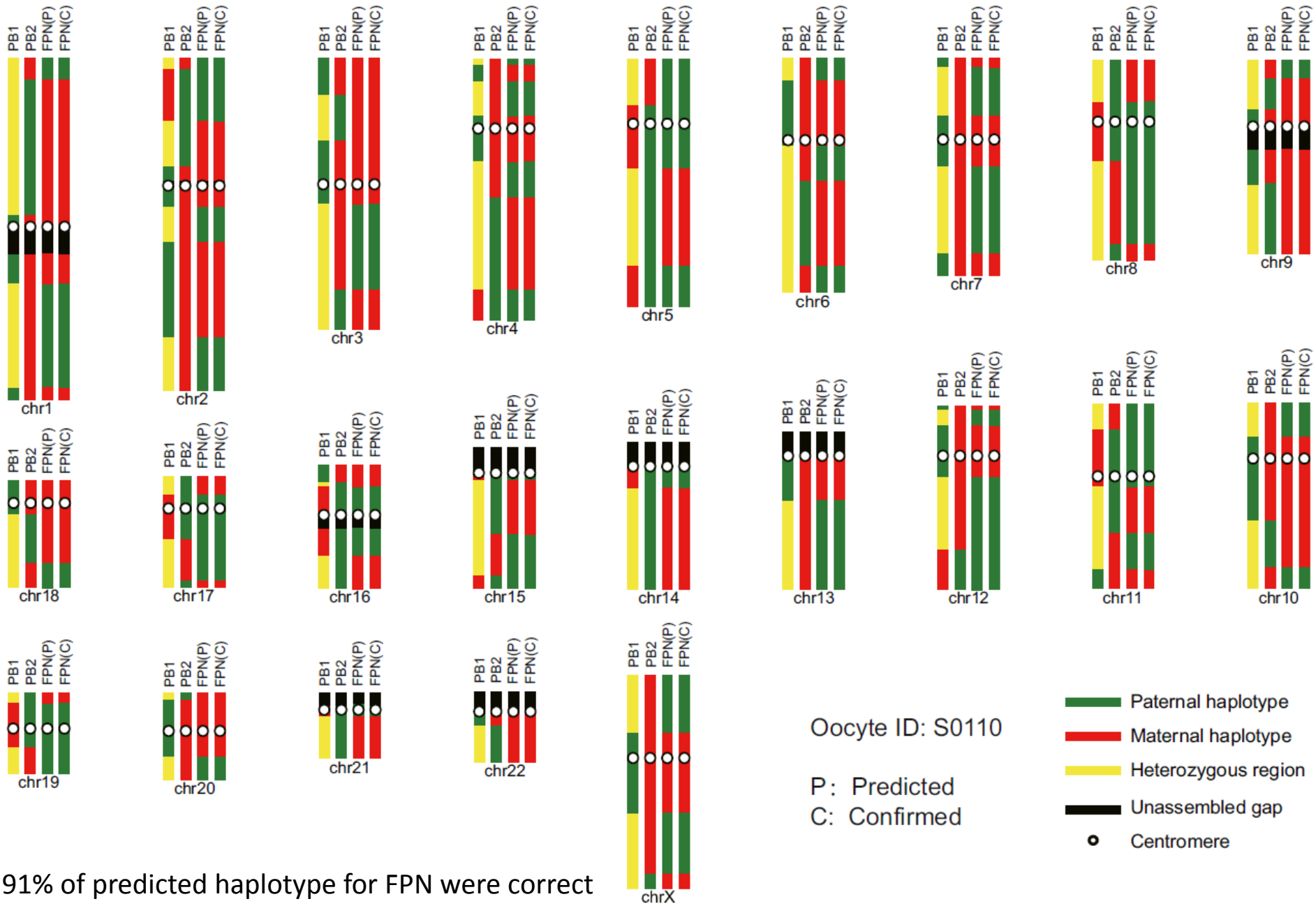


At each heterozygous locus of the donor:

- If PB1 is heterozygous, the haplotype of FPN should be different from that of PB2

- If PB1 is homozygous, the haplotype of FPN should be the same as that of PB2

Deduction of female pronucleus haplotype: proof-of-principle



91% of predicted haplotype for FPN were correct

Identification of maternal Mendelian diseases

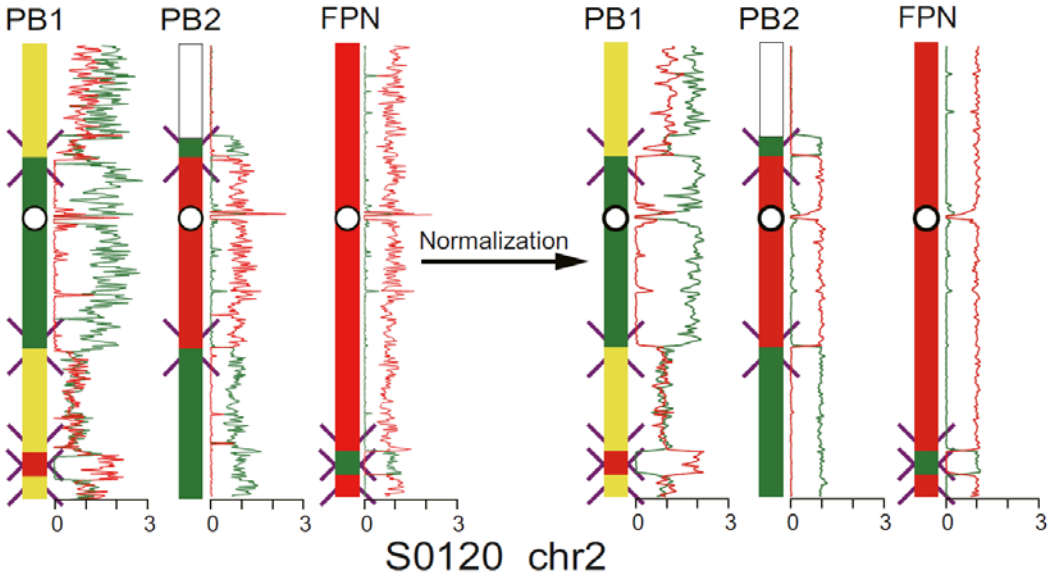
Table 1. Deduction of Mendelian Disease-Associated SNVs in Female Pronuclei of Donor S08

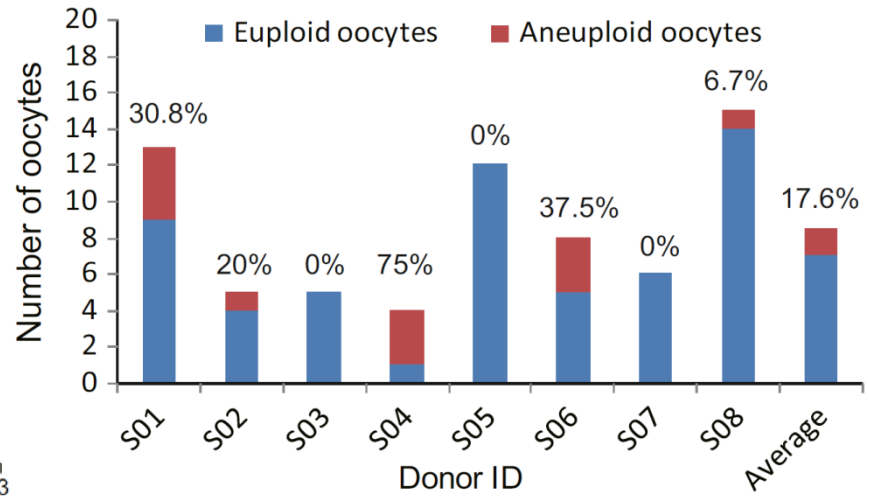
Oocyte ID	Gene Name											
	AGL (C→T)				*GP9* (G→A)				*HPD* (C→T)			
	PB1	PB2	FPN (P)	FPN (C)	PB1	PB2	FPN (P)	FPN (C)	PB1	PB2	FPN (P)	FPN (C)
S0801	N/N	N	T	T	N/N	N	G	G	T/N	C	C	C
S0802	N/N	N	C	C	N/N	N	G	G	N/N	T	T	T
S0803	N/N	N	T	T	N/N	N	G	G	N/N	C	C	C
S0806	N/N	C	C	C	N/N	N	A	A	N/N	N	C	C
S0807	N/N	N	C	C	N/N	N	A	A	N/N	N	C	C
S0808	C/N	T	C	C	N/N	N	A	A	N/N	N	C	C
S0809	C/C	N	T	T	N/N	N	G	G	N/N	T	C	C
S0811	N/N	T	T	T	N/N	N	A	A	N/N	N	C	C
S0812	C/C	T	T	T	N/N	N	G	Δ	C/T	N	T	T
S0813	N/N	C	T	T	N/N	N	G	G	N/N	T	C	C
S0815	N/N	N	C	C	N/N	N	A	A	N/N	N	T	T

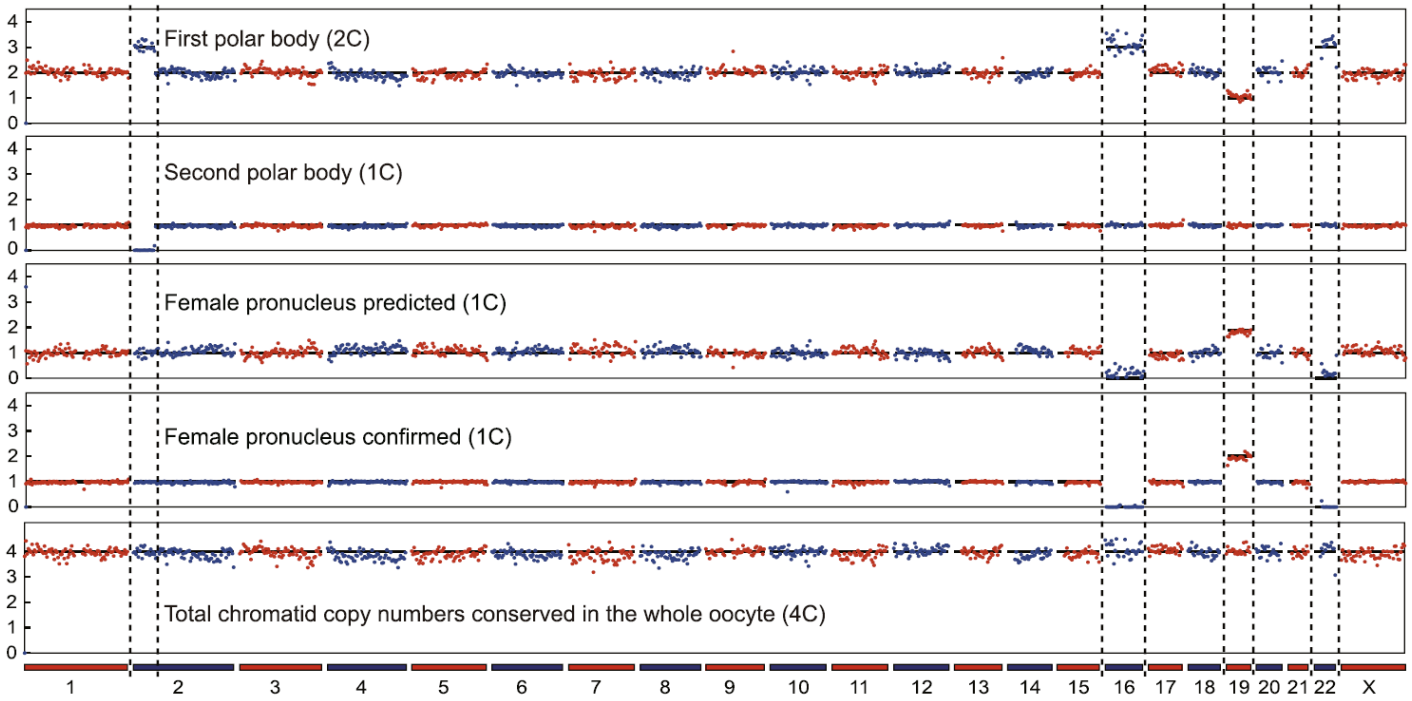
For this donor, the corresponding genotypes for 11 FPNs of the oocytes, which have all the PB1, PB2, and FPN recovered and sequenced, are predicted by the haplotype deduction. The "N" represents the alleles that are not covered by single-cell low-depth sequencing. The underlined bases represent SNVs associated with Mendelian diseases. Δ represents the allele that could not be amplified by locus-specific PCR from single-cell MALBAC product. Although some SNP loci are not covered by low-depth sequencing in polar bodies, the genotype of FPN can still be accurately predicted by the haplotype information. See also Figure S5 and Table S5.

- Screen for undesirable alleles predicted to be in female pronucleus
- Predicted alleles were confirmed by Sanger sequencing of the MALBAC product

High-resolution and accurate deduction of aneuploidies







MALBAC for oocytes – summary and future improvements

Current achievements:

- First comprehensive analysis of female meiosis recombination (included detailed analysis of chromosome and chromatide interference)

- Proof-of-principle application for prenatal genetic screen in *in vitro* fertilization for the unbiased identification of aneuploidies and disease-associated alleles without touching the growing embryo

Current limitations and future improvements:

- Paternal-inherited abnormalities are not detected
    - The same procedure can be applied to blastocysts

- Expensive procedure
    - Sequencing and analysis costs are expected to further decrese

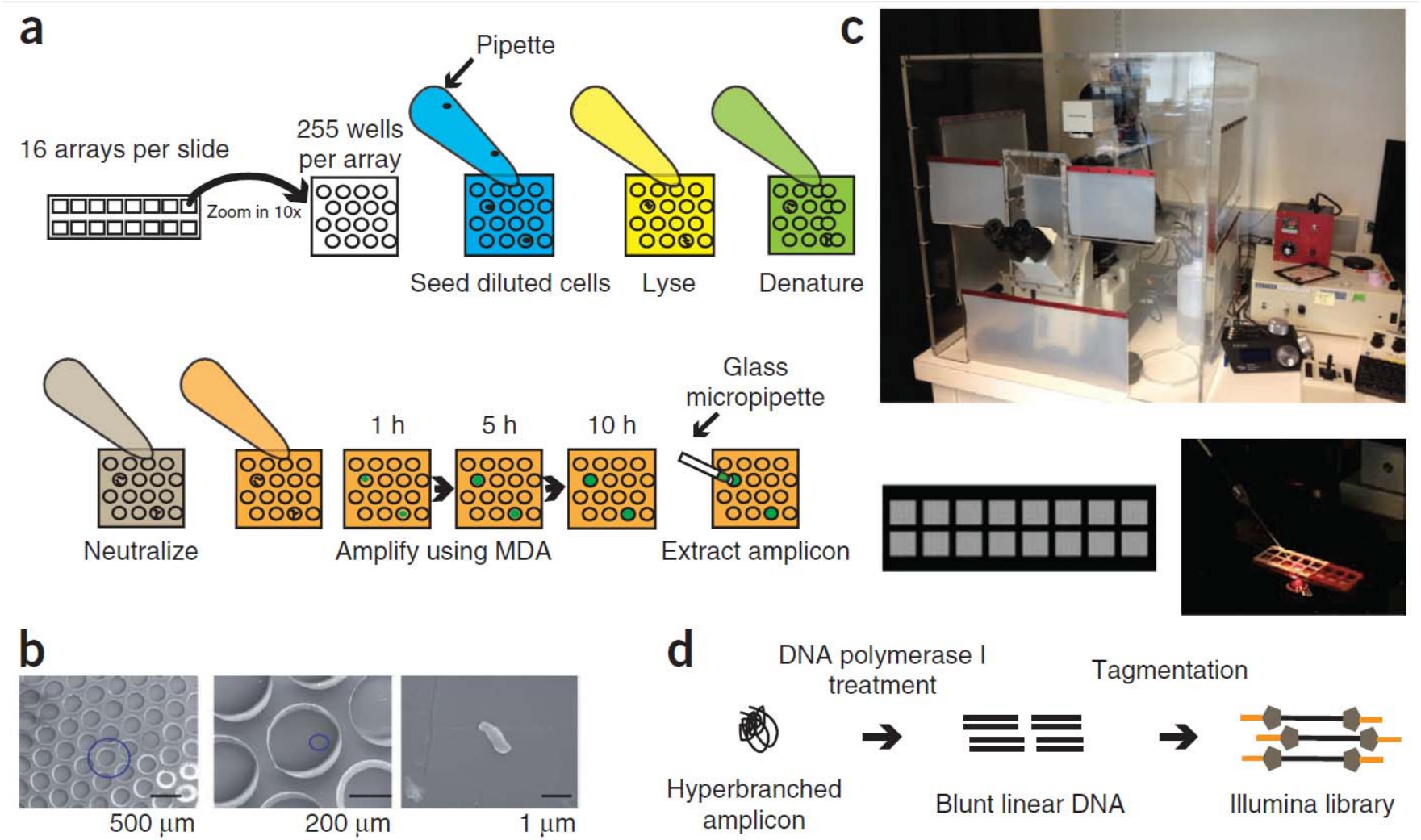
VOLUME 31 NUMBER 12 DECEMBER 2013 NATURE BIOTECHNOLOGY

nature biotechnology

Massively parallel polymerase cloning and genome sequencing of single cells using nanoliter microwells

Jeff Gole[1], Athurva Gore[1], Andrew Richards[1], Yu-Jui Chiu[2], Ho-Lim Fung[1], Diane Bushman[3], Hsin-I Chiang[1,5], Jerold Chun[3], Yu-Hwa Lo[4] & Kun Zhang[1]

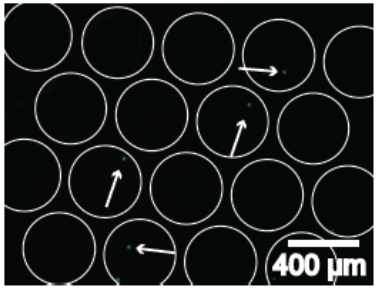
Microwell Displacement Amplification System: MIDAS

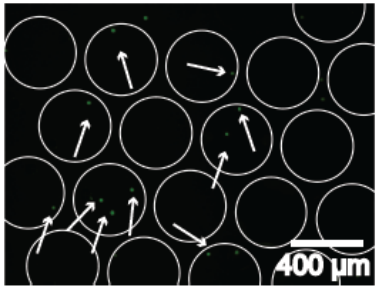


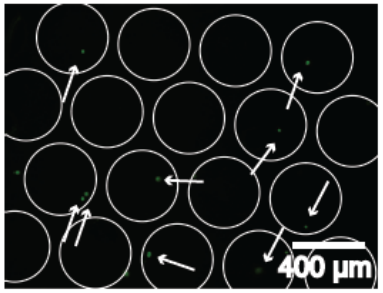
- All liquid handling procedures (cell seeding, lysis, DNA denaturation, neutralization and addition of amplification master mix) require one pump of a pipette per step per array
- Working volume: 12 nl/well

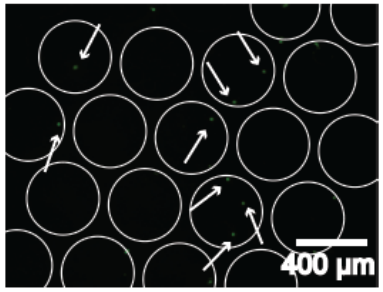
MIDAS: seeding

1 cell/well seeeding:
Most wells should have 1 cell
26% of wells could have more than 1 cell

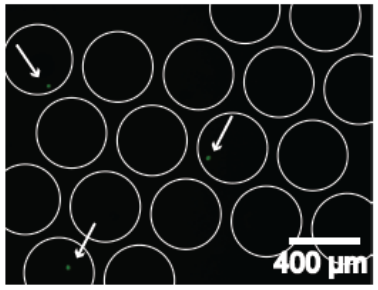


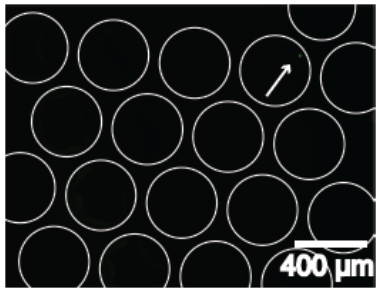


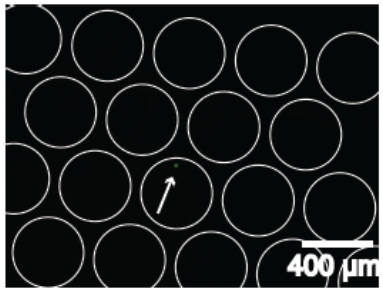


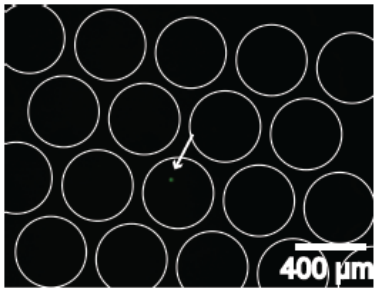


0.1 cell/well seeeding:
Most well should be empty
0.5% of wells could have more than 1 cell



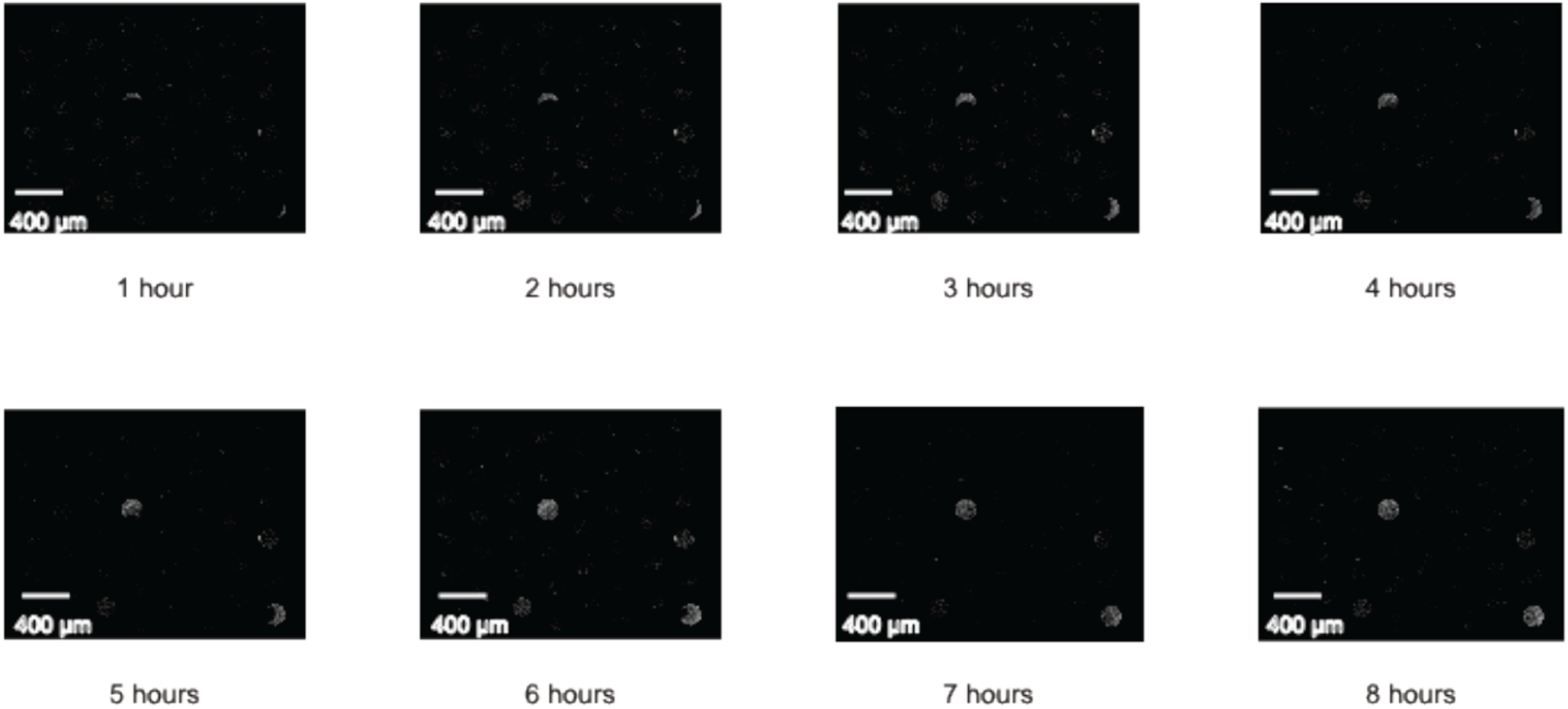






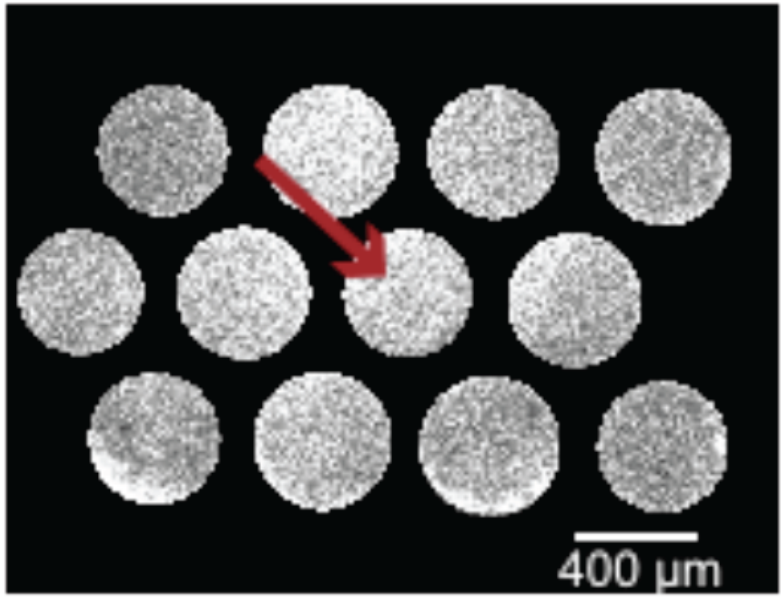
→ 0.1 cell/well seeding used for subsequent analyses

MIDAS: real time DNA amplification

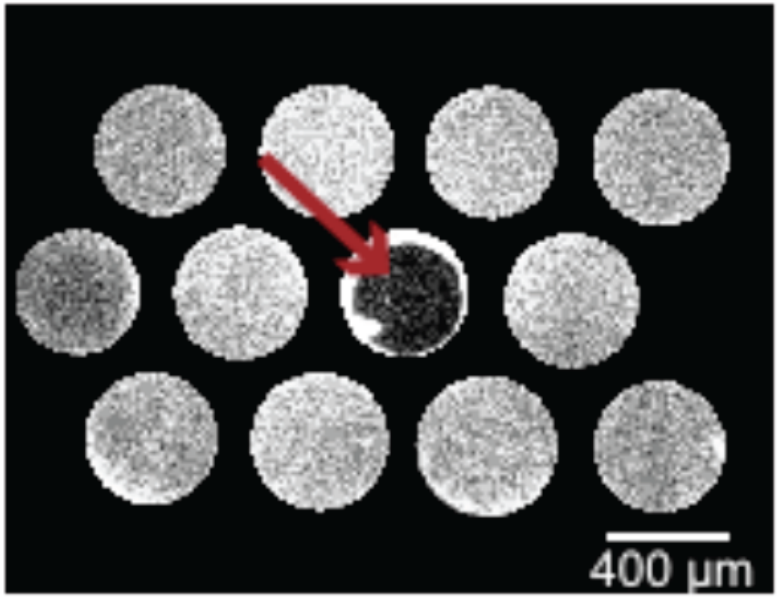


- SYBR Green I-labeled amplicons followed in real-time
- Continuing growing until saturation (after 5-6 h)
- Amplicons are randomly distributed, in line with random seeding
- No amplicons in abutting wells, excluding inter-well contamination

MIDAS: amplicon extraction



Before Extraction

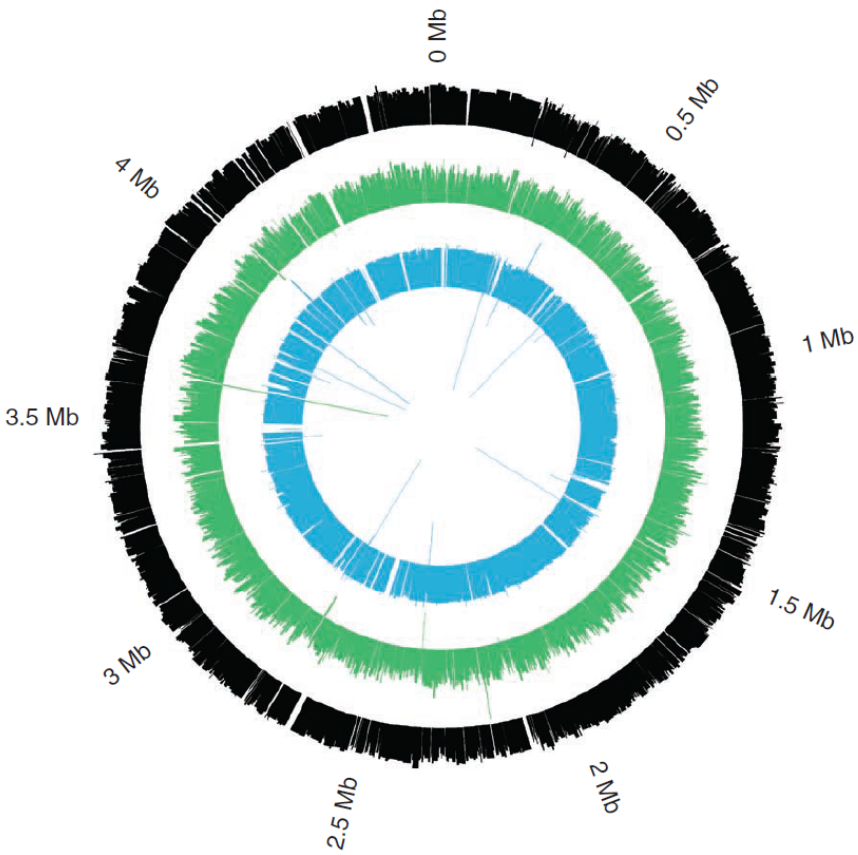


After Extraction

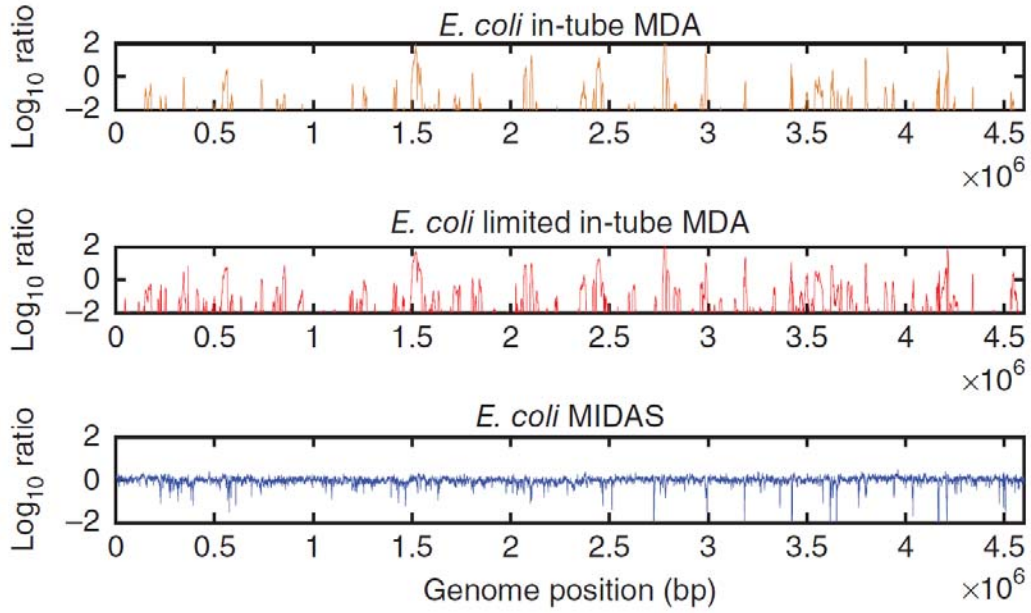
- Successful removal of desired amplicon with micropipette
- Performed manually (≈10 extraction/hour)

Generation of a near-complete assembly from single *E. coli*

MIDAS sequencing of 3 individual *E.coli* cells:



MIDAS vs MDA sequencing of individual *E.coli* cells:

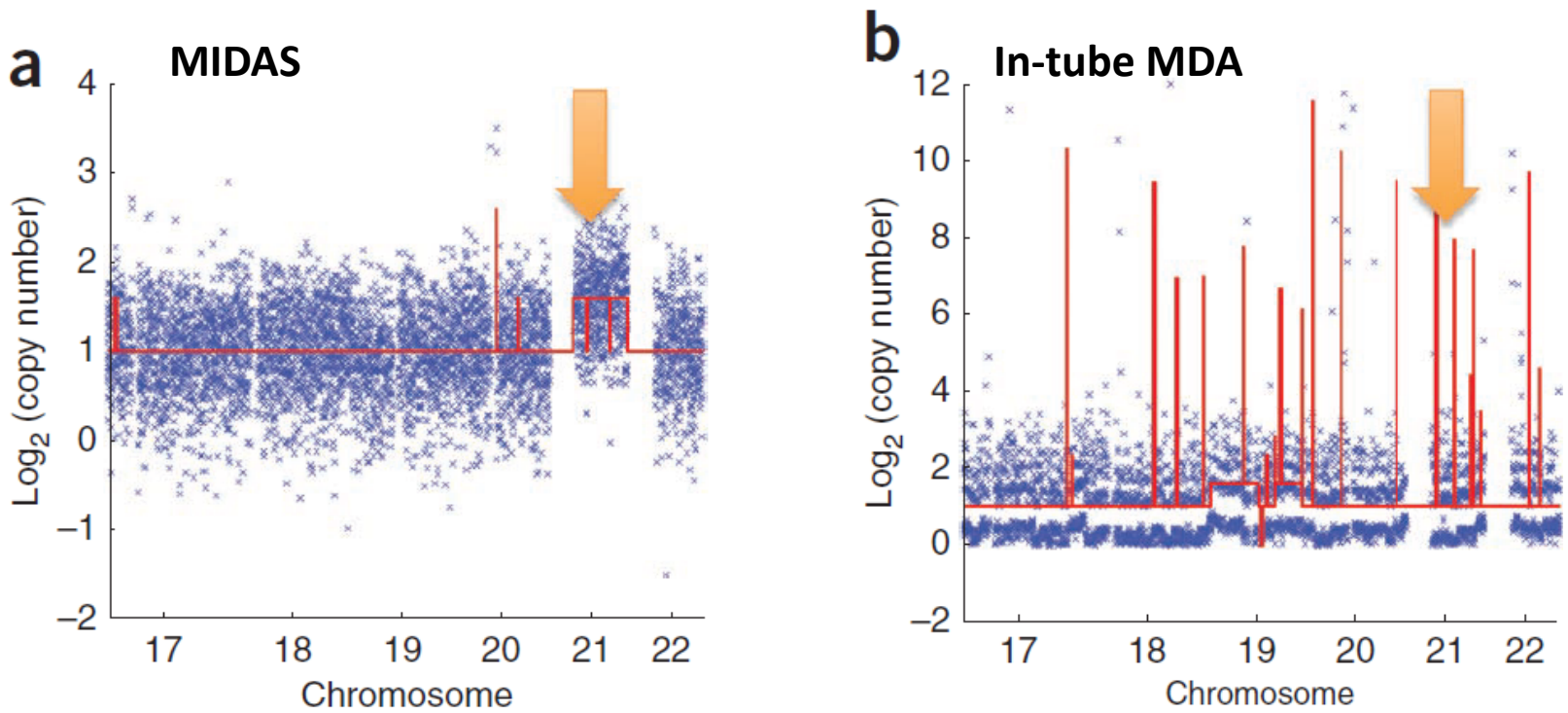


- 98-99% of the genome sequenced at >1x coverage
- ≈90% of the genome was correctly assembled *de novo*
- >96% E.coli genes correctly annotated

With respect to MDA:

- Lower amplification bias
- Higher genome coverage with less sequencing

Identification of copy number variations in single neurons



- Neuronal nuclei FACS-sorted from post-mortem brain of an invididual with Down syndrome and from a control subject

- MIDAS, but not in-tube MDA, could correctly identify Chr. 21 trisomy

MIDAS – summary and future improvements

Current achievements:

- Successful assembly of E.coli genome

- Successful identification of CNVs in neurons

Current limitations and future improvements:

- Only 10% of wells are exploited for amplification/sequencing due to low cell seeding
    - Increasing cell seeding density and implementing fluorescence monitoring of amplification to exclude wells with >1 cell or with cross-contamination from neighboring wells

- Access to microfabrication facility required
    - Commercial availability of hydrophilic microwell arrays

- Amplicon extraction performed manually
    - Implementing robotic automation for micropipetting

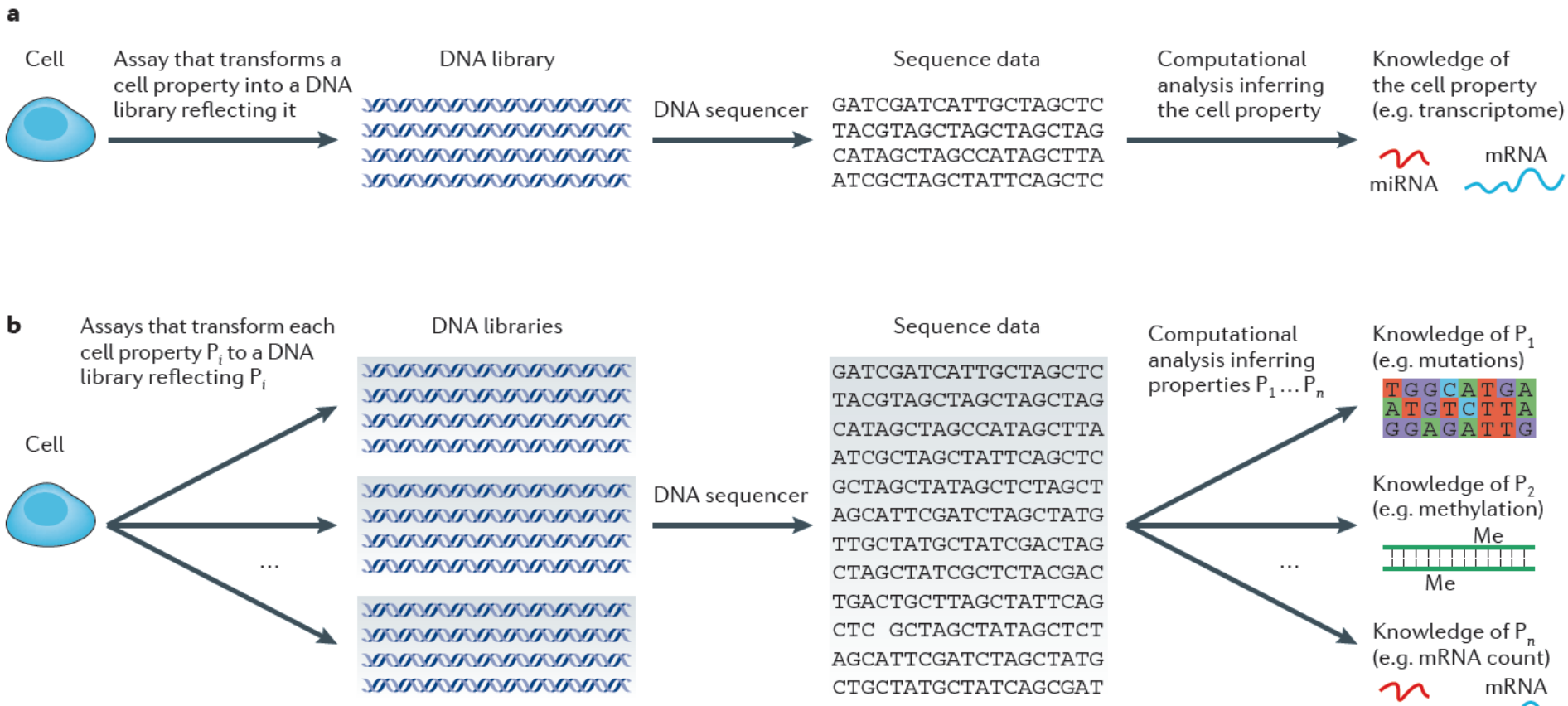
Single-cell genome sequencing: what next?

- Further refinements and wider applications

- Integration of:
  - Single-cell genomics
  - Single-cell transcriptomics
  - Single-cell epigenomics
  - Single-cell proteomics

APPLICATIONS OF NEXT-GENERATION SEQUENCING

Single-cell sequencing-based technologies will revolutionize whole-organism science



Shapiro et al. Nat Rev Genet 2013