Biological & Medical Science based on Physics:
Radiation and physics, Physics on medical science,
Modeling for biological system

Panasonic Hall at the Yukawa Memorial Building
Yukawa Institute for Theoretical Physics
Kyoto University:
November 5, 2015

Detection of spontaneous mutations in mammalian genomes

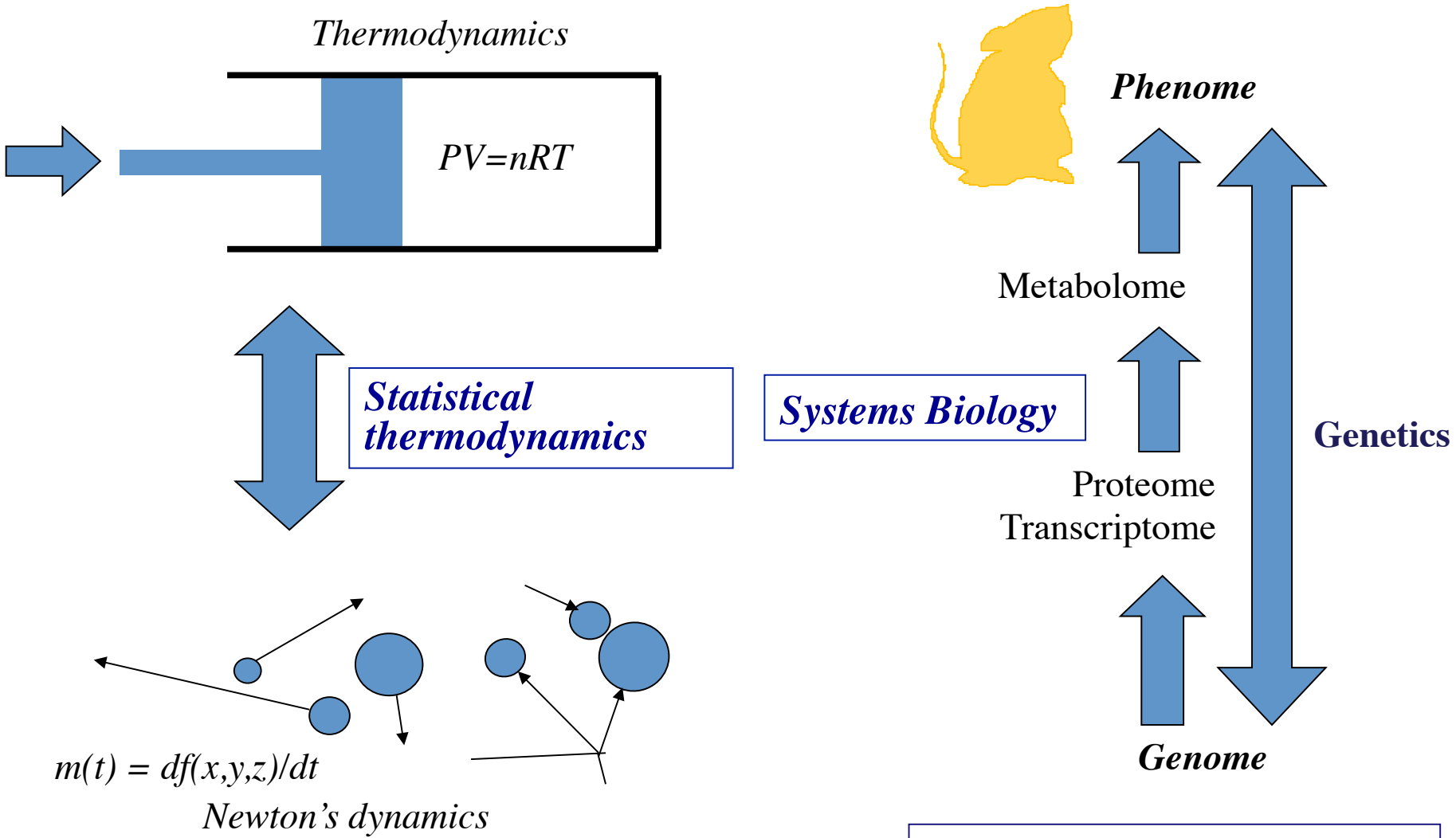
The 150-th Anniversary of Mendel's Discovery

Yoichi Gondo

Mutagenesis and Genomics Team

RIKEN BioResource Center

From Elements to Whole



Mutations Cause Human Genetic Diseases in Various Organs and Systems

Brain
Huntington's disease
Dentatorubral palidoluisyan atrophy
Fragile X syndrome
Alzheimer's disease
Perkinson's disease
Prader-Willi S/Angelman syndrome
Down syndrome

Spine
Spinocerebellar ataxia
Spinal and bulbar muscular atrophy
Spinal mascular atrophy
Amyotrophic lateral scleorosis

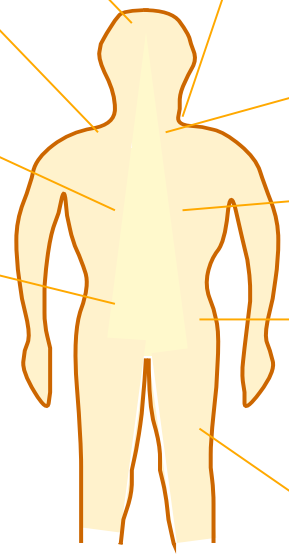
PNS
Charcot-Marie-Tooth disease
Familial amiloidotic polyneuropathy
Neurofibromasis

Skin
Xeroderma pigmentosum
Cockayne's syndrome

Eyes
Retinoblastoma
Achromatopsia

Blood
Tharathemia
Hemophiria

Tumor
ras
myc
p53



Thymus
Multiple endocrine neuroplasia

Lung
Cystic fibrosis

Brest&Ovary
BRCA

Heart
Hypertrophic cardiomyopathy

Kidney
Wilm's tumor type 1

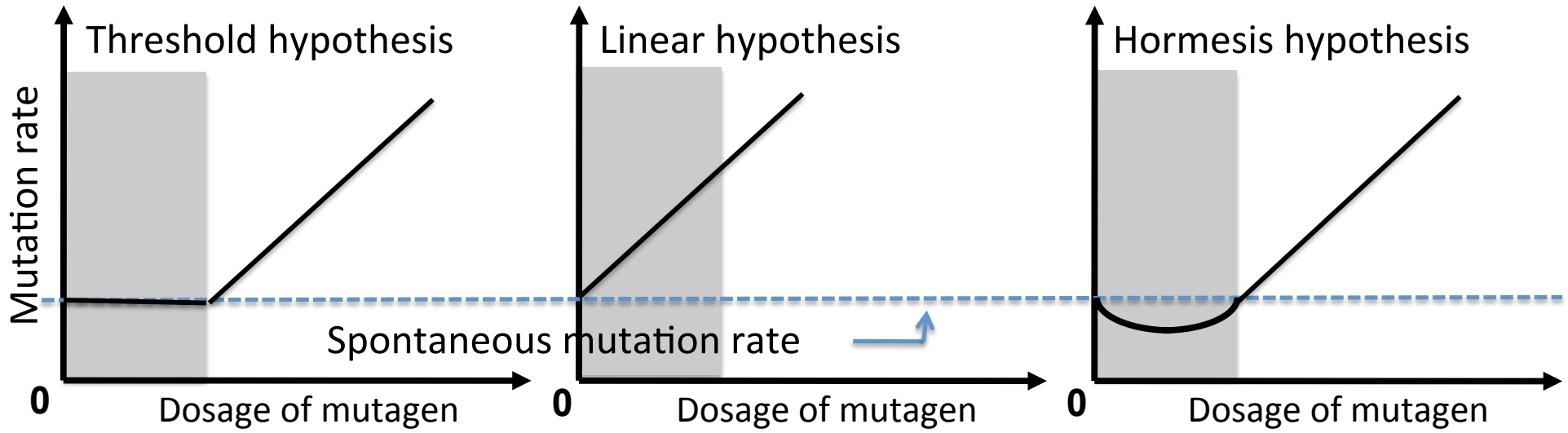
Colon
Familial adenomatous polyposis
Hereditary non-polyposis colorectal cancer

Metabolic
Diabetes mellitus
Hyperlipidemia
Phenylketonurea
Alkaptonurea

Muscle
Duchenne
Dystrophi

3×10^{9}bp/haploid
6×10^{13} cells/Adult
Total of 3.6×10^{23} bp of DNA/Adult
Mutation rate: $10^{-8} \sim 10^{-12}$ /bp/cell division

Question: How bad (or good) of the low dose environmental genotoxic factors (mutagens)?



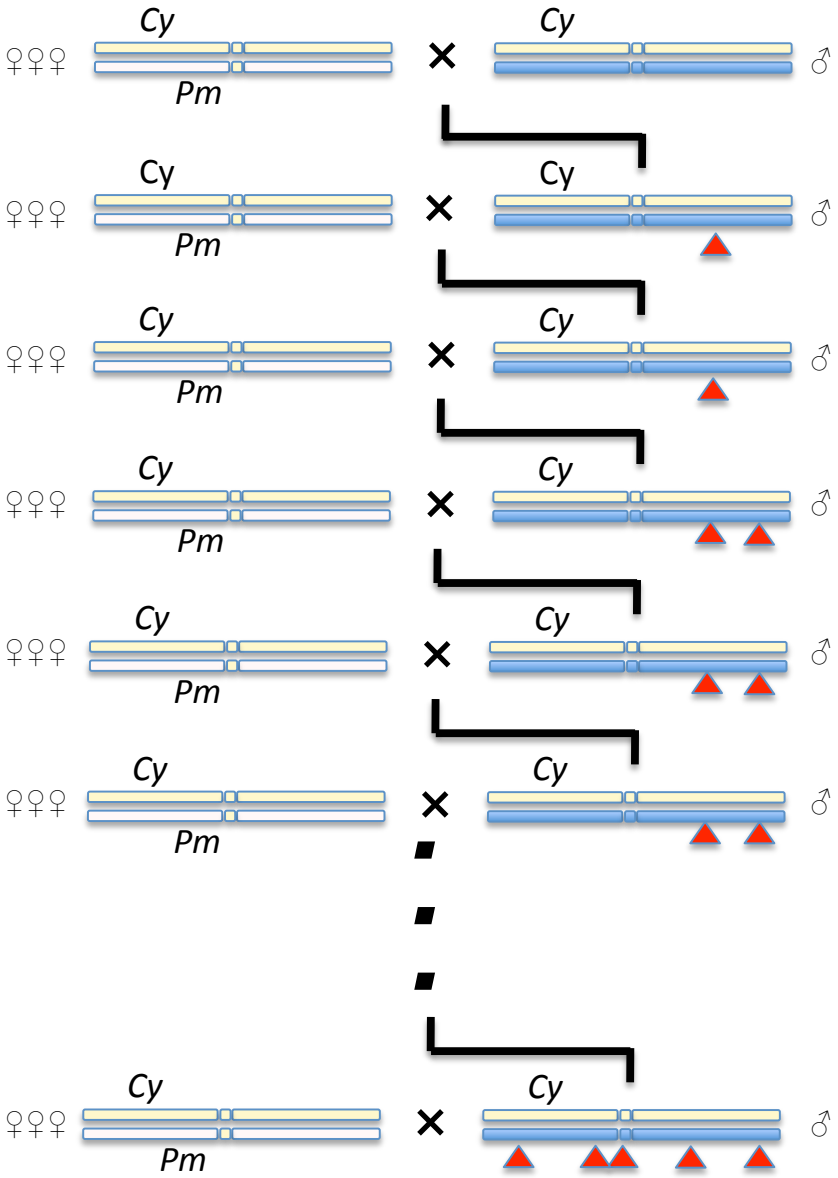
Cy/Pm method: Accumulation of Germline Mutations
-- *D. melanogaster* (fruit fly) --

Balancer chromosome

Cy

Carries a dominant visible marker and a recessive lethal mutation with multiple inversions that surpress genetic recombination.

.

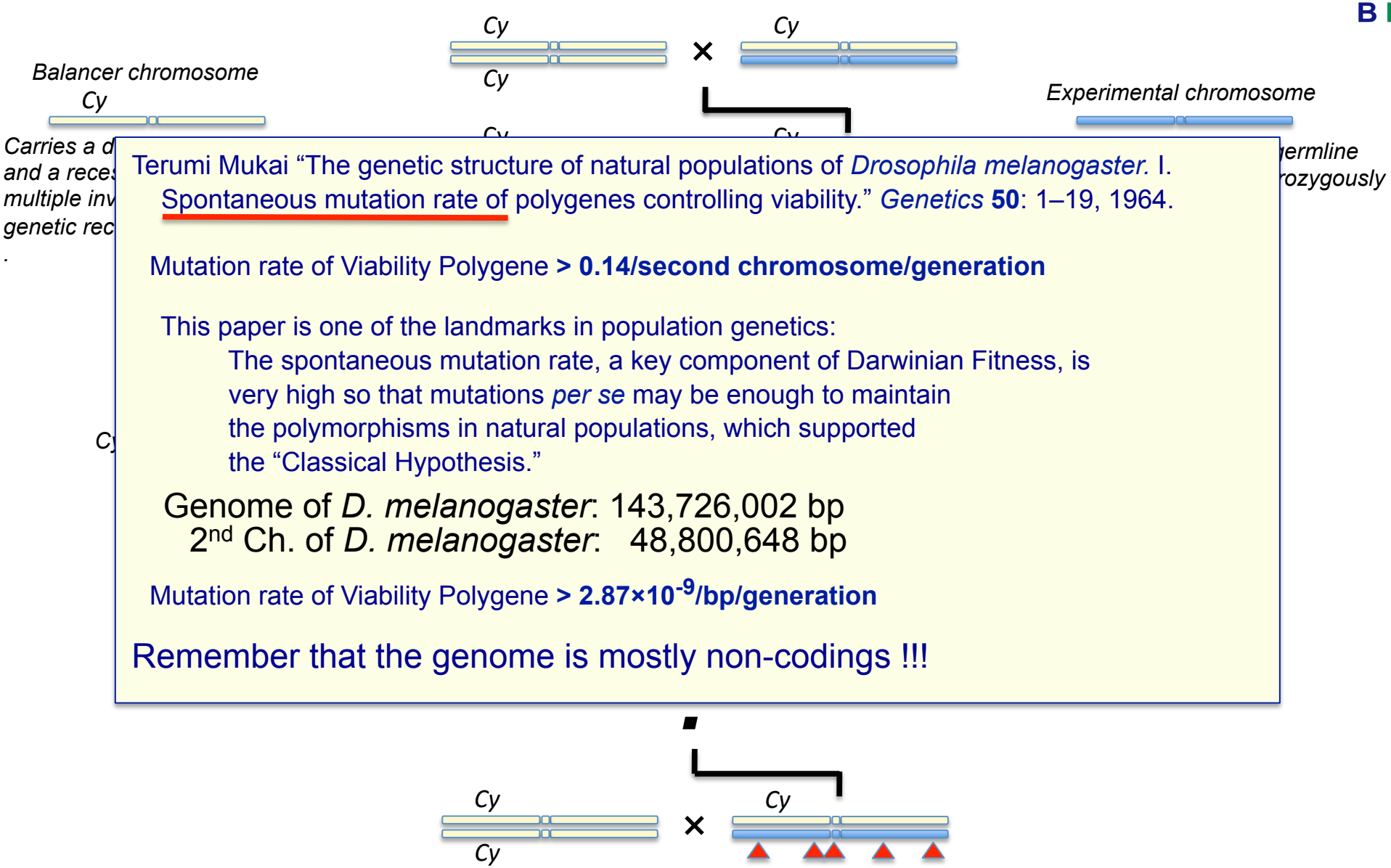


Experimental chromosome

on which de novo germline mutations are heterozygously accumulated.

Cy method: Accumulation of Germline Mutations

-- *D. melanogaster* (fruit fly) --



Terumi Mukai "The genetic structure of natural populations of *Drosophila melanogaster.* I. Spontaneous mutation rate of polygenes controlling viability." *Genetics* **50**: 1–19, 1964.

Mutation rate of Viability Polygene **> 0.14/second chromosome/generation**

This paper is one of the landmarks in population genetics:

The spontaneous mutation rate, a key component of Darwinian Fitness, is very high so that mutations *per se* may be enough to maintain the polymorphisms in natural populations, which supported the "Classical Hypothesis."

Genome of *D. melanogaster*: 143,726,002 bp
2nd Ch. of *D. melanogaster*: 48,800,648 bp

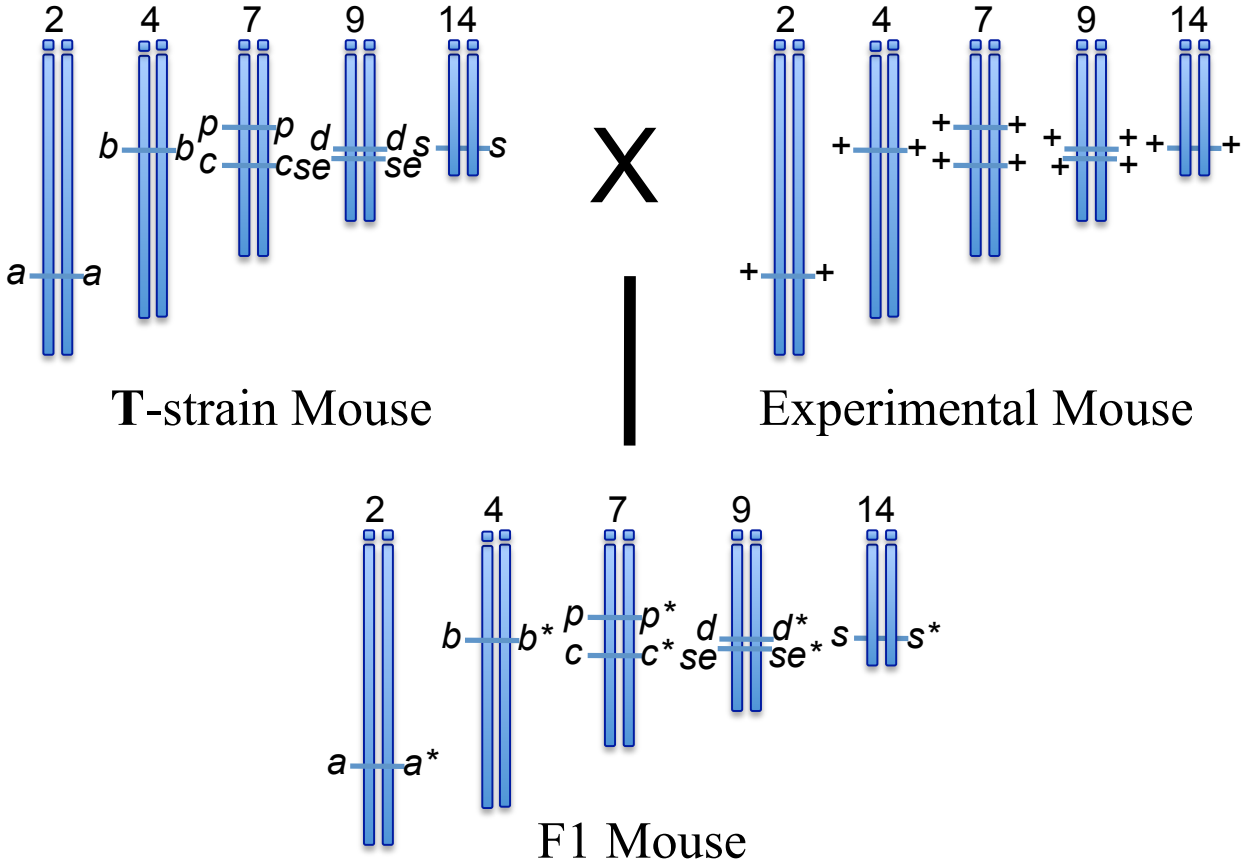
Mutation rate of Viability Polygene **> 2.87×10^{-9}/bp/generation**

Remember that the genome is mostly non-codings !!!

Detection of Mutations in the Mouse Genome

No "balancer chromosomes" are available in the mouse, basically.

Specific Locus Test by W. Russell

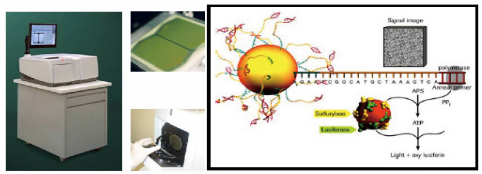


Spontaneous mutations rate was estimated as

~10^{-5}/locus/generation.

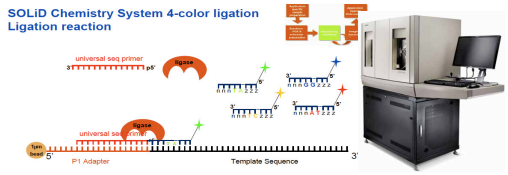
NGS systems

Roche454 GS-FLX



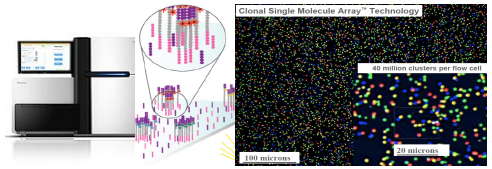
Read tag ~700bp
Total 0.7Gb (1M read tags/1 days)

AB SOLiD5500xl



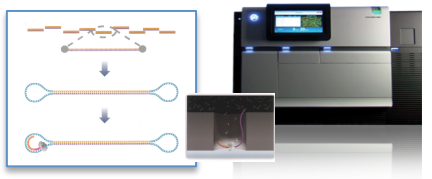
Read tag 60bp + 60bp
Total 180Gb (1.5G read tags/10 days)

Illumina HiSEQ2000



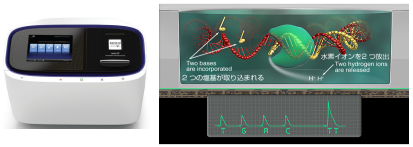
Read tag 100bp + 100bp
Total 600Gb (6G read tags/11 days)

PacBio RS



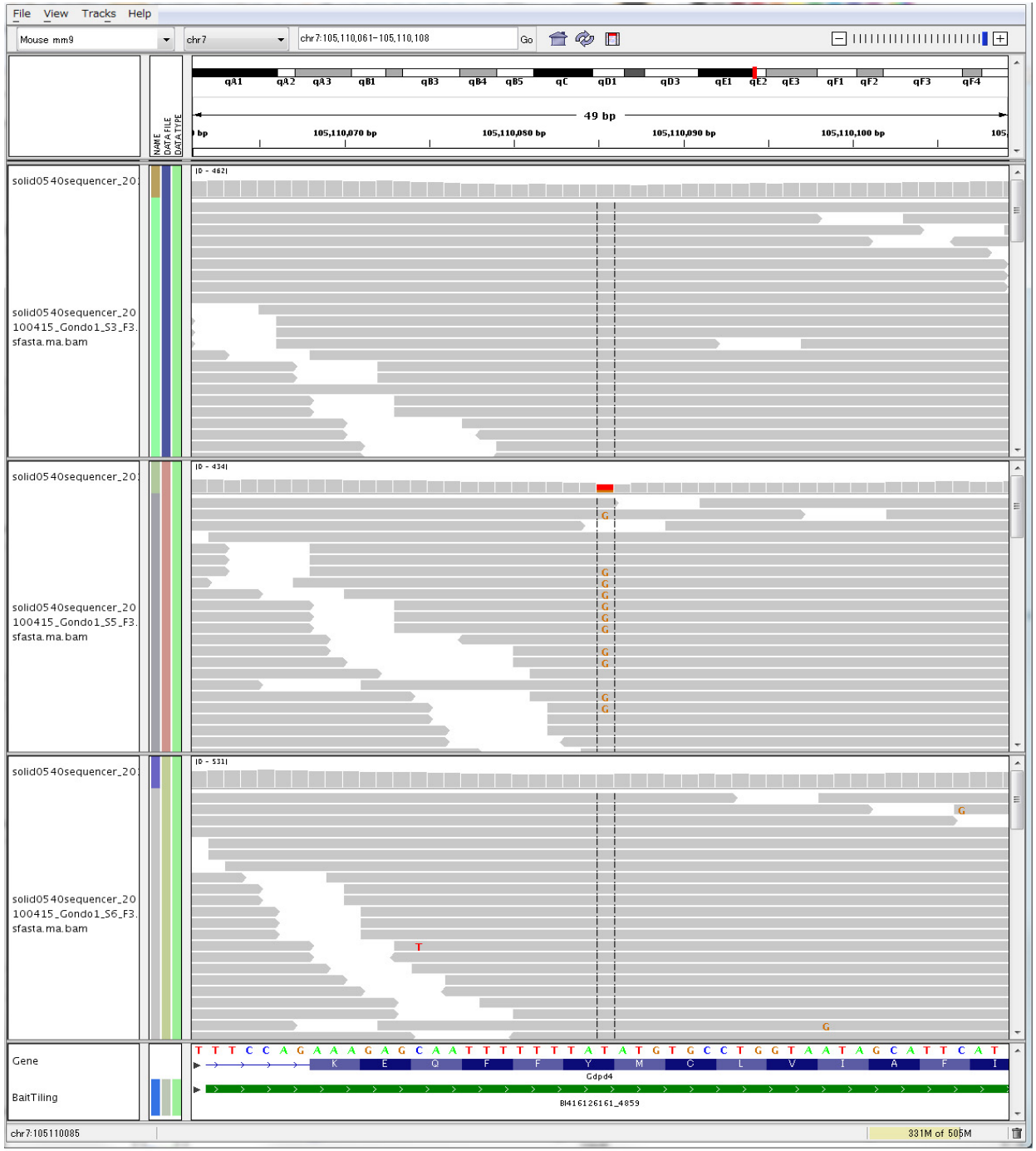
Read tag >3,000bp
Total 36Gb (12M read tags/1 days)

AB Ion Proton



Read tag ~200bp
Total 12Gb:Chip I (70M read tags/1 day)
30Gb:Chip II (200M read tags/1 day)

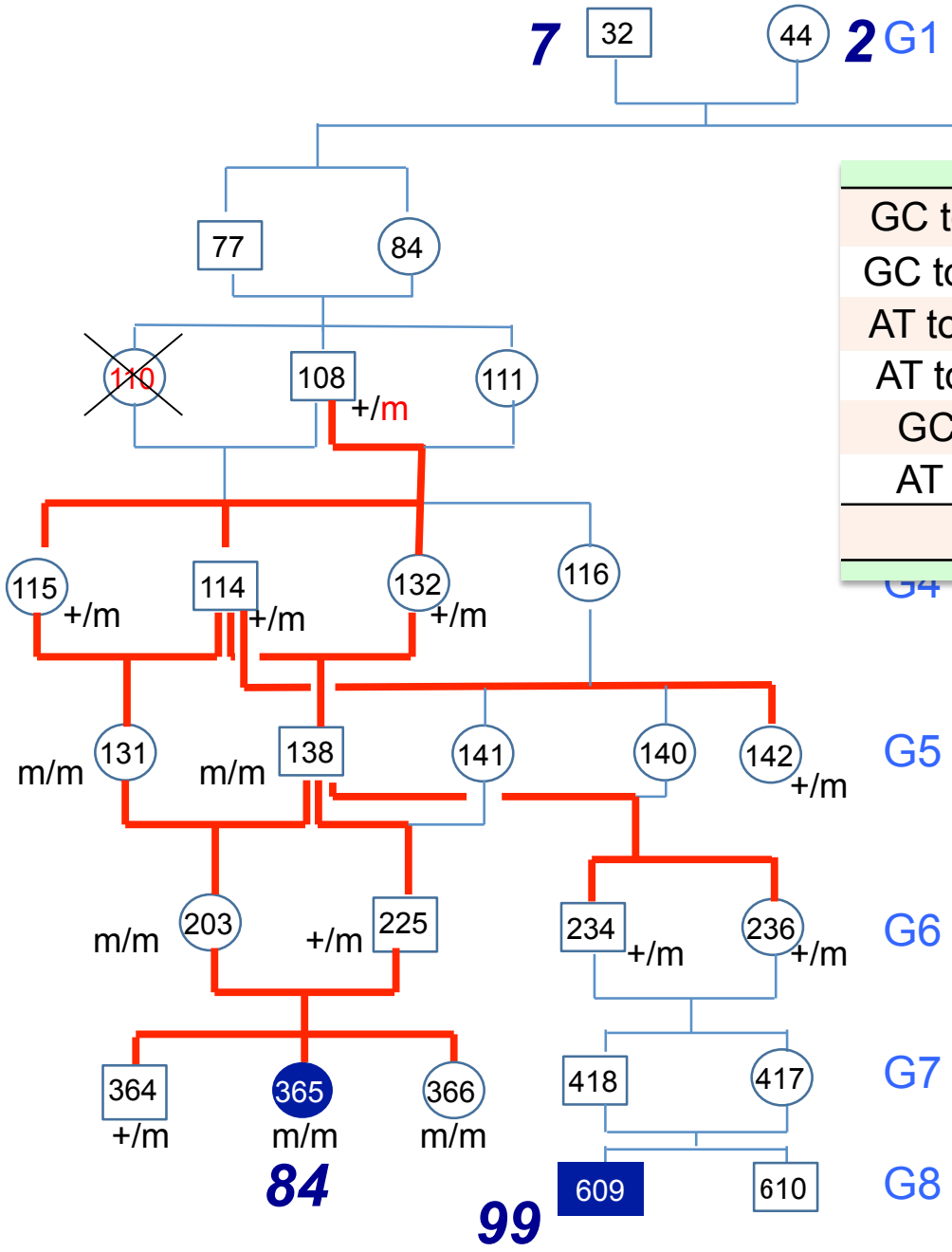
IGV browser allows “wet” scientists to see the NGS raw data.



Mutation rates estimated by NGS

Species	Mutation Rate		Ref
Human	12	$\times 10^{-9}$/bp/generation	Kong et al. 2012
Chimp	12	$\times 10^{-9}$/bp/generation	Venn et al. 2014
Mouse	5.4	$\times 10^{-9}$/bp/generation	Uchimura et al. 2015
Drosophila	2.8	$\times 10^{-9}$/bp/generation	Keightley et al. 2014
C. elegans	2.7	$\times 10^{-9}$/bp/generation	Denver et al. 2009

EXAMPLE OF WES application: Triple KO Pedigree Lacking Oxidative Stress Repairs



GC to TA transversions	252
GC to CG transversions	0
AT to CG transversions	1
AT to TA transversions	0
GC to AT transitions	7
AT to GC transitions	2
Sum	262

We were also able to depict the spectrum of mutations.

Sum = 297

We were able to trace all the identified mutations upto the origin by the MassArray system.

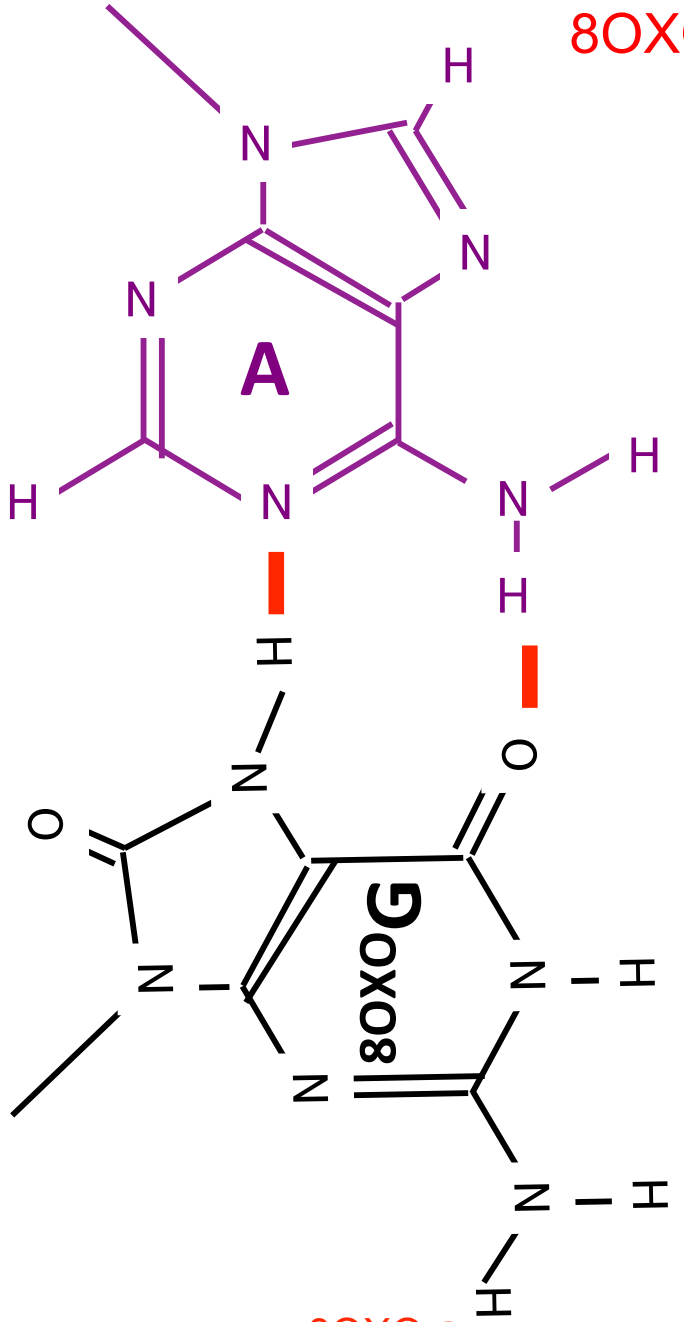
male

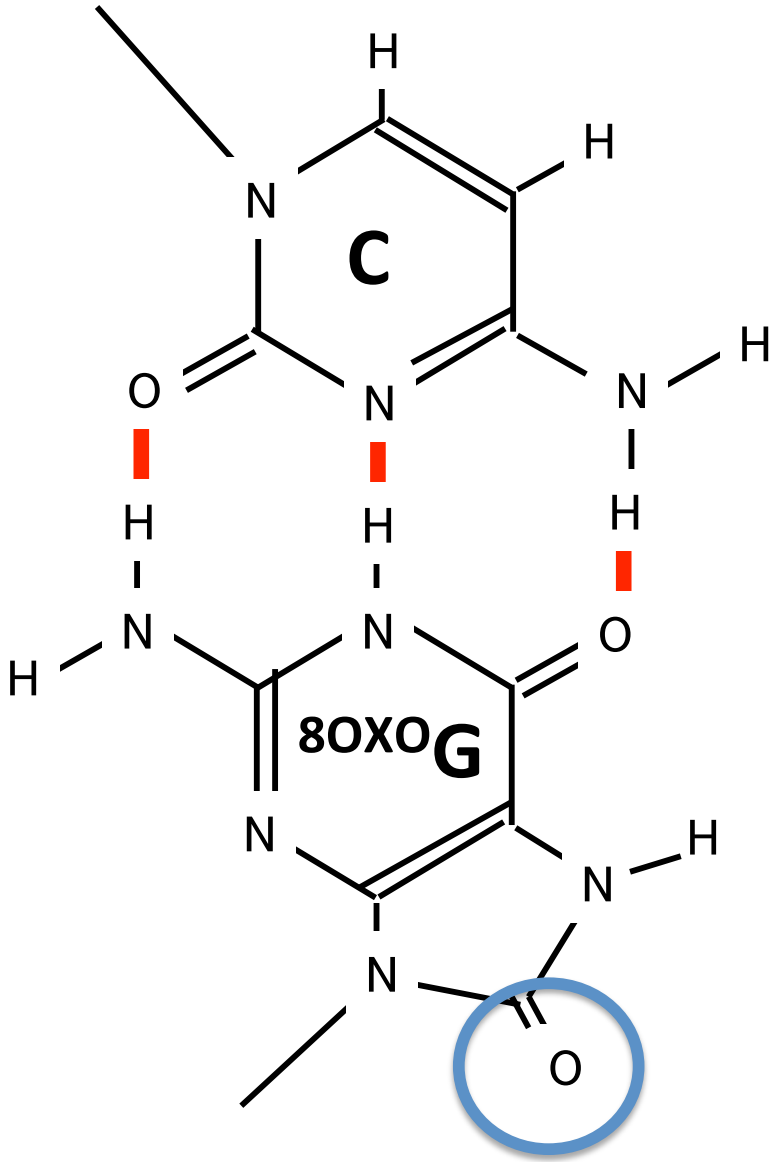
female

Three mice that were subjected to the exome sequencing with NGS.

Ohno et al. ***Scientific Reports*** 4: 4689, 2014.

^{8OXO}**G** may also pair with **A**

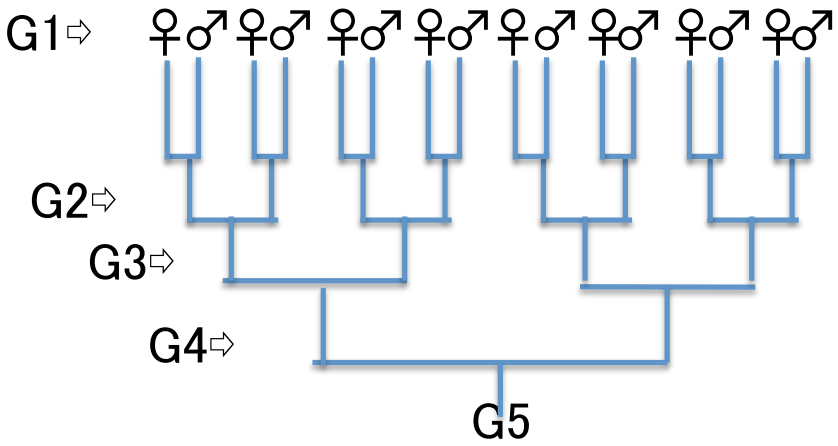




Consequently, ^{8OXO}**G**:**A** mismatch causes GC to TA transversion.

13

Universal Scheme to detect *de novo* spontaneous germline mutations



× 8 sets

Expected Mutation Rate: $m = 1 \times 10^{-8}$/bp/gamete/generation
Size of Mouse Genome: $S = 3 \times 10^{9}$ bp/gamete
Expected # of *de novo* Mutations: $n = m \times S = 30$/gamete/generation
Let
the predisposed # of mutation at $G_i = 2K_i$/diploid,
Then, the # of mutations at $G_{i+1} = (K_i + n) + (K_i + n) = (2K_i + 2n)$/diploid
Thus, the # of mutations at G5 = $2K_1 + 4 \times (2n) = 2K_1 + 8 \times 30 = 2K_1 + 240$

Total # of *de novo* mutation = 240 × 8 G5 = 1,920

*This method and knowledge are applicable to **QUALITY CONTROL** of any bioresource with sexual reproduction and RISK ASSESSMENT for the low-dose effects of various environmental mutagens.*

Supported by JSPS KAKENHI 2013~2017 to Y.Gondo

Preliminary Single Nucleotide Variation (SNV) Call of 23,800 From 4 G5 mouse genomes

Whole Genome Sequencing (WGS) of 8 G5 genomes completed.
- HiSEQ2500, rapid mode, 4 samples/17 lanes × 2 runs
- Pair End Read of ~450bp fragments
- 151 cycle
- Basecall: RTA1.17.21.3

SNV call for the beginning half of G5 genomes (= 4 G1 genomes) at present.
Mapping by BWA to mm10 mouse genome reference sequences.
Samtools to call SNVs.
Eliminate all homozygous SNVs as well as any common SNVs in 2 or more samples because such SNVs must have derived from the G1 mice and are not *de novo* mutations.

From the above pipeline, we obtained 23,800 SNV candidates.

Validation of TOP 113 SNV candidates out of the 23,800 Primary SNV call

G5 sample ID	C001-02	C001-03	C001-06	C001-08	Sum	R
Validated SNV candidates	**31**	**26**	**25**	**31**	**113**	**100%**
G1 Predisposed SNV	27	23	19	27	96	85%
de novo SNV	3	3	6	4	16	14%
N.D.	1*	0	0	0	1	1%

*This SNV was confirmed in G2 - G5 but the G1 data could not obiteined.

Predisposed SNV : de novo SNV ＝96:16＝ 6 : 1

Let

the predisposed # of mutation at G_i = $2K_i$/diploid,

Then, the # of mutations at G_{i+1} = $(K_i + n) + (K_i + n) = (2K_i + 2n)$/diploid

Thus, the # of mutations at G5 = $2K_1 + 4 \times (2n) = 2K_1 + 8 \times 30 = 2K_1 + 240$

Predispoed SNV : de novo SNV in one G5 genome

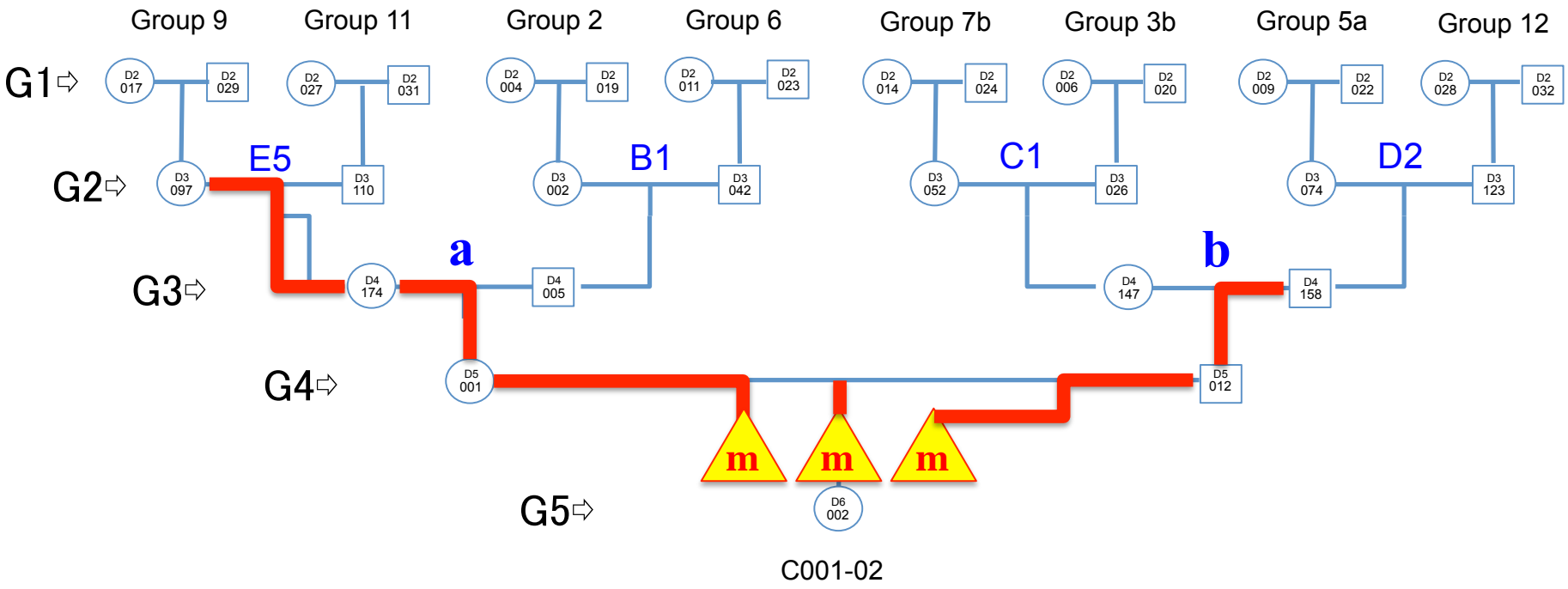
$= 2K_i + 240 = 1{,}440 + 240$

Validation of TOP 113 SNV candidates out of the 23,800 Primary SNV call

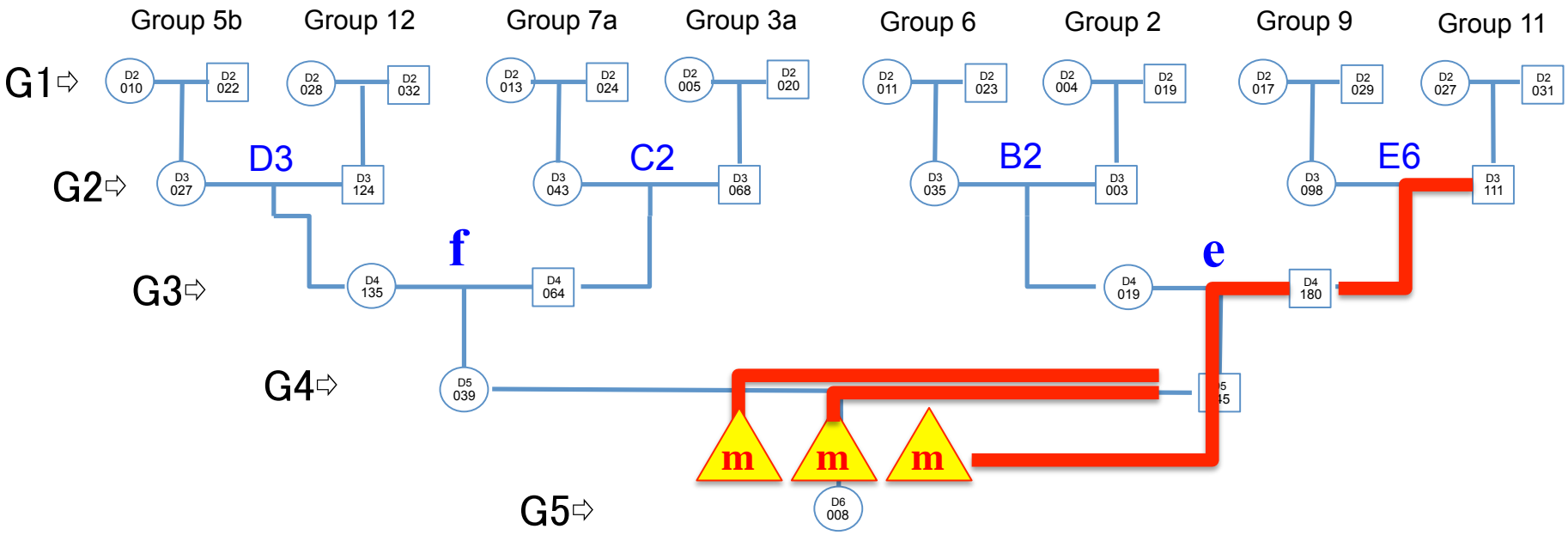
G5 sample ID	C001-02	C001-03	C001-06	C001-08	Sum	R
Validated SNV candidates	**31**	**26**	**25**	**31**	**113**	**100%**
G1 Predisposed SNV	27	23	19	27	96	85%
de novo SNV	3	3	6	4	16	14%
N.D.	1*	0	0	0	1	1%

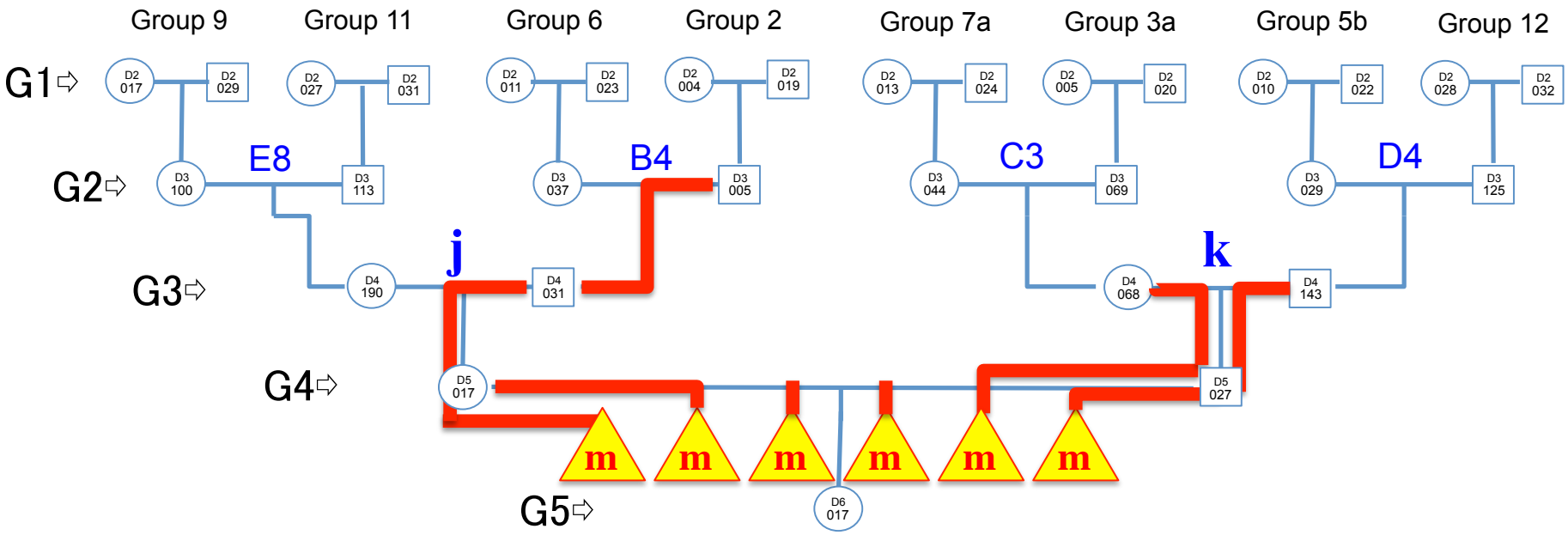
*This SNV was confirmed in G2 - G5 but the G1 data could not obiteined.

de novo SNV	3	3	6	4	16	100%
G2 Origin	1	1	1	1	4	25%
G3 Origin	1	0	2	1	4	25%
G4 Origin	0	2	1	2	5	31%
G5 Origin	1	0	2	0	3	19%

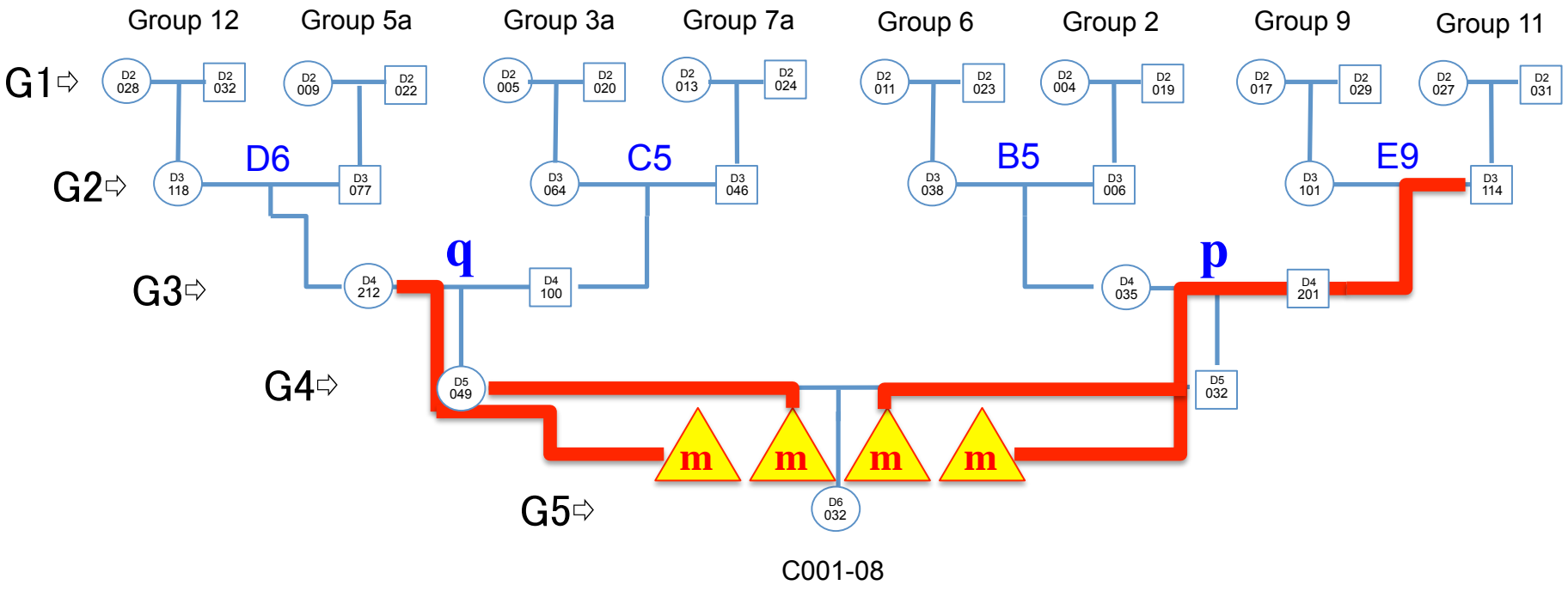

Group 9
Group 11
Group 2
Group 6
Group 7b
Group 3b
Group 5a
Group 12
G1
G2
G3
G4
G5
D2 017
D2 029
D2 027
D2 031
D2 004
D2 019
D2 011
D2 023
D2 014
D2 024
D2 006
D2 020
D2 009
D2 022
D2 028
D2 032
D3 097
D3 110
D3 002
D3 042
D3 052
D3 026
D3 074
D3 123
E5
B1
C1
D2
D4 174
D4 005
D4 147
D4 158
a
b
D5 001
D5 012
m
m
m
D6 002

C001-02


Group 5b
Group 12
Group 7a
Group 3a
Group 6
Group 2
Group 9
Group 11
G1
G2
G3
G4
G5
D2 010
D2 022
D2 028
D2 032
D2 013
D2 024
D2 005
D2 020
D2 011
D2 023
D2 004
D2 019
D2 017
D2 029
D2 027
D2 031
D3 027
D3 124
D3 043
D3 068
D3 035
D3 003
D3 098
D3 111
D4 135
D4 064
D4 019
D4 180
D5 039
D6 008
D3
C2
B2
E6
f
e
m
m
m
C001-03


Group 9
Group 11
Group 6
Group 2
Group 7a
Group 3a
Group 5b
Group 12
G1
G2
G3
G4
G5
D2 017
D2 029
D2 027
D2 031
D2 011
D2 023
D2 004
D2 019
D2 013
D2 024
D2 005
D2 020
D2 010
D2 022
D2 028
D2 032
D3 100
D3 113
D3 037
D3 005
D3 044
D3 069
D3 029
D3 125
E8
B4
C3
D4
D4 190
D4 031
D4 068
D4 143
j
k
D5 017
D5 027
m
m
m
m
m
m
D6 017

C001-06


Group 12
Group 5a
Group 3a
Group 7a
Group 6
Group 2
Group 9
Group 11
G1
G2
G3
G4
G5
D2 028
D2 032
D2 009
D2 022
D2 005
D2 020
D2 013
D2 024
D2 011
D2 023
D2 004
D2 019
D2 017
D2 029
D2 027
D2 031
D3 118
D6
D3 077
D3 064
C5
D3 046
D3 038
B5
D3 006
D3 101
E9
D3 114
D4 212
q
D4 100
D4 035
p
D4 201
D5 049
m
m
m
m
D5 032
D6 032
C001-08

800 Primary SNV call

	C001-08	Sum	R
Va	31	113	100%
G1	27	96	85%
	4	16	14%
	0	1	1%

data could not obiteined.

6 : 1

Let

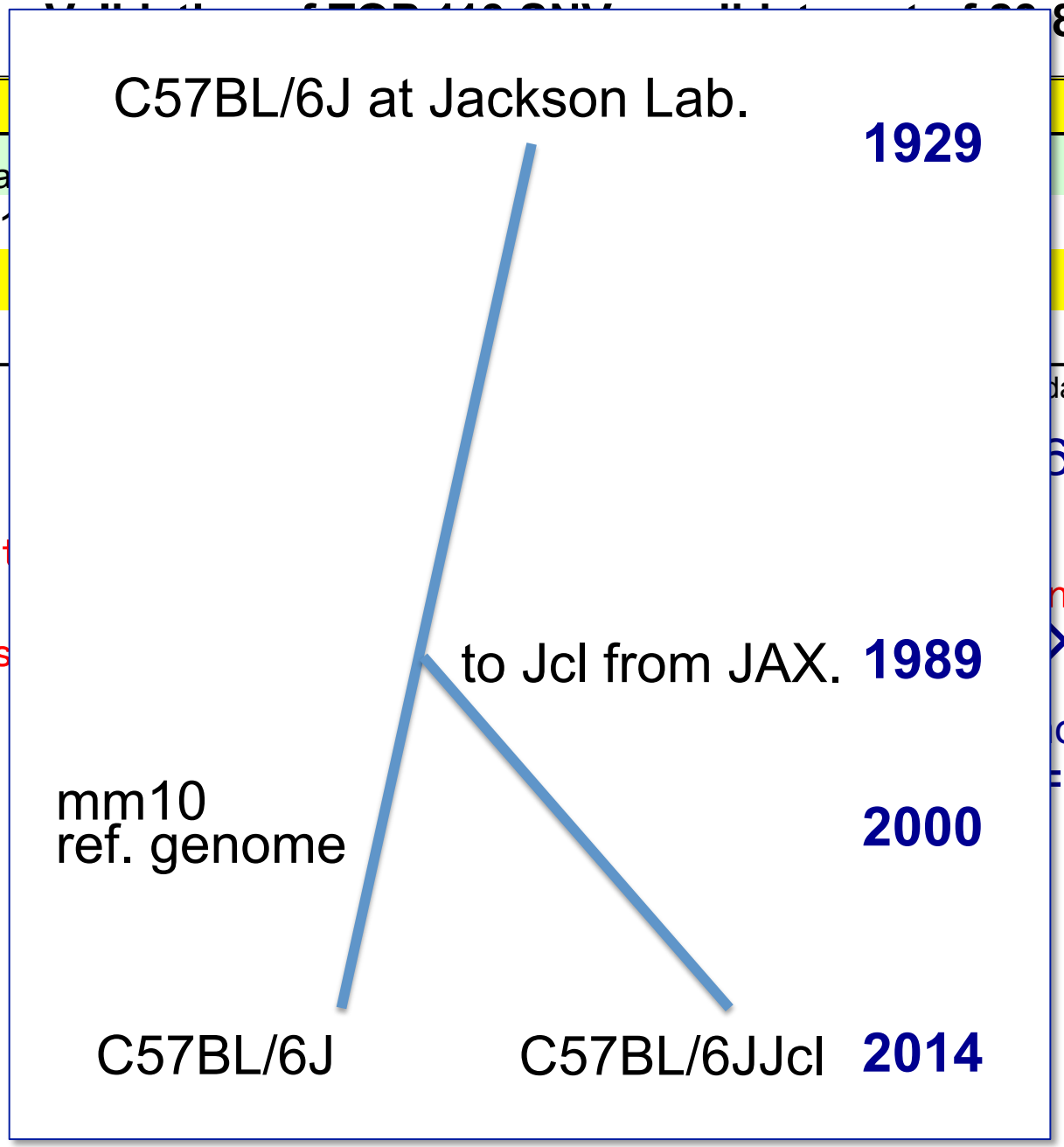
Thus

n) = (2Ki + 2n)/diploid

× 30 = 2K1 + 240

ome

= 1,440 + 240



SUMMARY

We developed a universal mutation accumulation scheme for any sexually reproductive species.

Whole genome sequencing (WGS) with new generation sequencers (NGS) effectively detected *de novo* as well as predisposed single nucleotide variations (SNVs).

Quick and accurate genotyping systems (*e.g.*, MassArray, AmpliSeq) allow us to trace and pinpoint the origin of SNVs.

Detected *de novo* SNVs give the mutation rate while both *de novo* and predisposed SNVs provide the data for spectrum studies.

Even in the mouse, ~1,000 *de novo* spontaneous germline SNVs could be accumulated in a year, which is “large-scale” enough and quick enough to depict the background mutation rate for the assessment of low-dose mutagen effects, *e.g.*, low-dose irradiation, chemical environmental risk factors, etc.

This scheme also becomes a tool to conduct studies on “experimental” molecular evolution.

Almost all of SNVs, if not all, may be mildly deleterious mutations affecting viability polygenes.

CONTRIBUTORS

RIKEN BioResource Center

Mutagenesis and Genomics Team

Ryutaro Fukumura
Shigeru Makino
Yuji Nakai
Hayato Kotaki
Yuichi Ishitsuka
Hideko Nemoto
Masako Tsuruga

National Institute of Genetics

Comparative Genomics Laboratory

Atsushi Toyoda
Asao Fujiyama

Kyushu University

Dept. of Biosci. and Biotec., Fac. Agr.

Kazuki Mori
Satoru Kuhara

Medical Institute of Bioregulation

Mizuki Ohno
Kunihiko Sakumi
Yusaku Nakabeppu

Special Thanks to

Dr. Terumi Mukai

Dr. Tomoko Ohta

Dr. Akiyoshi Wada

This work is partly supported by JSPS KAKENHI No. 15200032, 21240043, 25241016, MEXT KAKENHI 211S0002 and AMED.