



SNP Microarray Interpretation Guide



Timothy Fee, Ph.D.
Greenwood Genetic Center
Greenwood, South Carolina
Version 1 - April 2021

Table of Contents

Page

Introduction.....	3
Description of Included Cases	4
Simulated SNP Microarray Data	5
<i>Single Aberrations</i>	
Single Copy Gain	5
Single Copy Loss	6
Two Copy Gain.....	7
Two Copy Loss.....	8
Loss of Heterozygosity (LOH).....	9
Single Copy Gain with Three Haplotypes	10
Parental Cell Contamination	11
<i>Combinations of Two Aberrations</i>	
Single Copy Gain + Single Copy Loss	12
Single Copy Gain + LOH	13
Single Copy Gain + Two Copy Gain	14
Single Copy Gain + Two Copy Loss	15
Single Copy Gain + Gain with Three Haplotypes	16
Single Copy Gain + Parental Cell Contamination.....	17
Single Copy Loss + LOH	18
Single Copy Loss + Two Copy Gain	19
Single Copy Loss + Two Copy Loss	20
Single Copy Loss + Gain with Three Haplotypes	21
Single Copy Loss + Parental Cell Contamination.....	22
LOH + Two Copy Gain	23
LOH + Two Copy Loss	24
LOH + Gain with Three Haplotypes	25
LOH + Parental Cell Contamination.....	26
Two Copy Gain + Two Copy Loss	27
Two Copy Gain + Gain with Three Haplotypes	28
Two Copy Gain + Parental Cell Contamination.....	29
Two Copy Loss + Gain with Three Haplotypes	30
Two Copy Loss +Parental Cell Contamination.....	31
Gain with Three Haplotypes + Parental Cell Contamination.....	32

Alternatively, refer to the table below to identify the page number for a specific combination of aberrations:

	Gain	Loss	LOH	Two-Copy Gain	Two-Copy Loss	Gain w/ Three Haplotypes	Parental Cell Contamination
Gain	5	12	13	14	15	16	17
Loss	12	6	18	19	20	21	22
LOH	13	18	7	23	24	25	26
Two Copy Gain	14	19	23	8	27	28	29
Two-Copy Loss	15	20	24	27	9	30	31
Gain w/Three Haplotypes	16	21	25	28	30	10	32
Parental Cell Contamination	17	22	26	29	31	32	11

Introduction

Single-nucleotide polymorphism (SNP) microarrays have found significant clinical utility over the past decade for the detection of both constitutional and somatic cytogenetic abnormalities. SNP arrays contain probes that identify the genotype of a SNPs throughout the genome. The advantage of having SNP probes in addition to copy-number probes is the ability to a) detect regions of homozygosity, b) corroborate copy-number probe data, and c) determine the haplotype of individuals. However, the pattern of SNP data can be non-intuitive for complex or mosaic situations. Therefore, an *in silico* model of a SNP array was developed to generate simulated SNP array data. The simulated array results can be used as a tool for learning SNP array analysis or as a reference for interpretation of SNP array data. This document contains simulated SNP array data which can be used as an educational resource for anyone interested in the analysis of SNP microarray data.

The *in silico* SNP array model used to generate this data is essentially a Monte Carlo simulation. The basic steps are as follows:

- 1) A large number of virtual cells are generated
- 2) In a subset of these cells, a specific abnormality is induced
- 3) For each simulated SNP, the number of times each allele is observed is counted
- 4) The allele difference and B-Allele Frequency (BAF) values are computed
- 5) The allele difference values or BAF values are plotted similar to how SNP array data is displayed

The following formulas are used to compute the allele difference and BAF values.

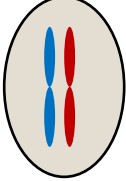
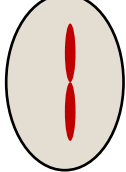
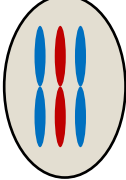
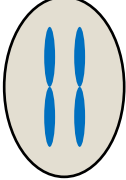
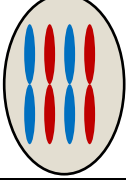
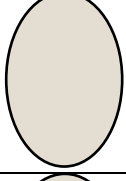
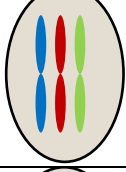
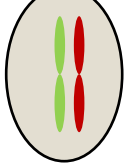
$$\text{Allele Difference} = \frac{(\text{Number of A Alleles}) - (\text{Number of B Alleles})}{2 \times \text{Number of Cells}}$$

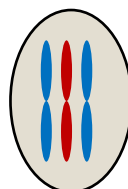
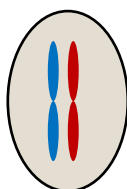
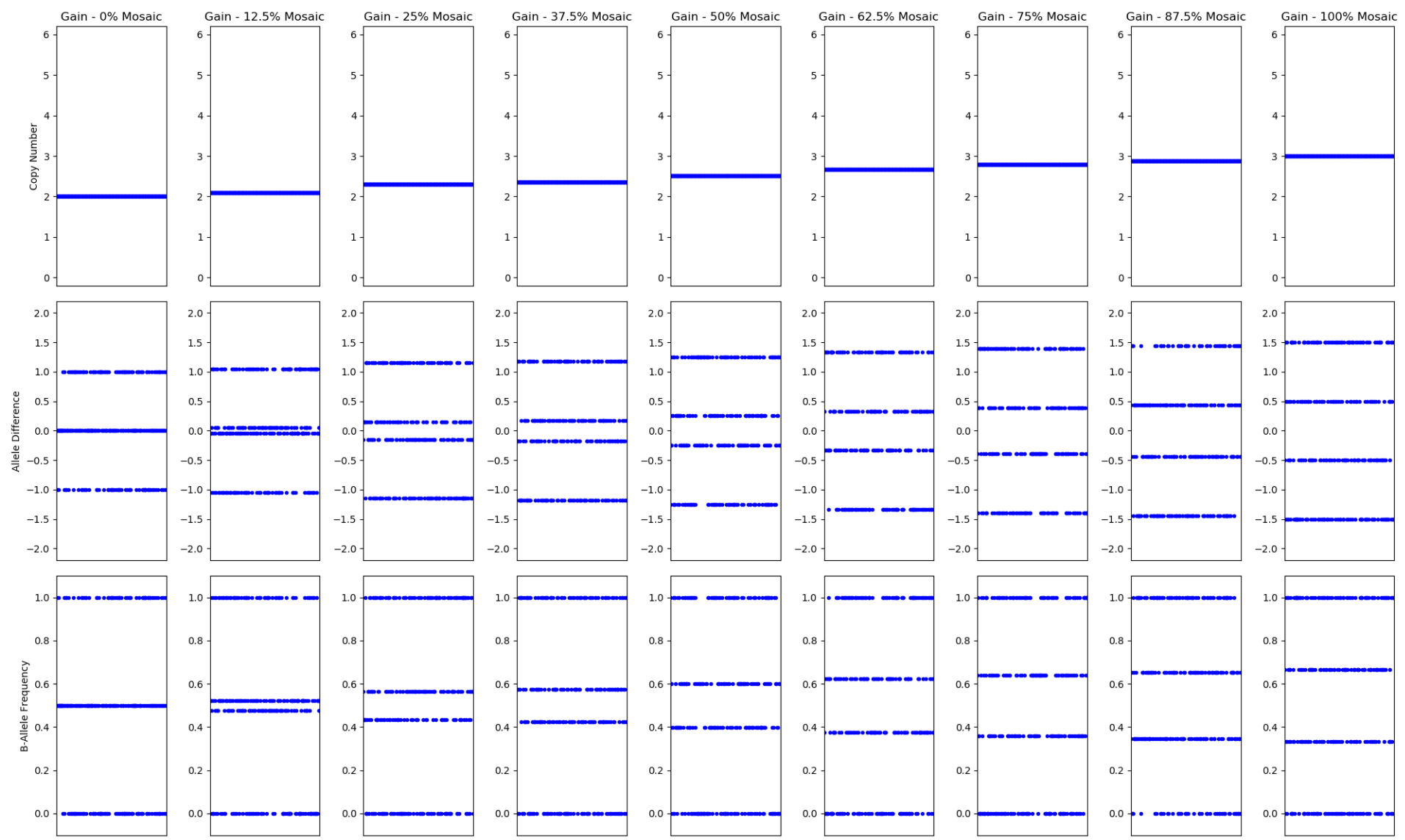
$$\text{BAF} = \frac{\text{Number of B Alleles}}{(\text{Number of A Alleles}) + (\text{Number of B Alleles})}$$

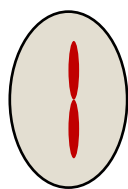
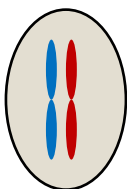
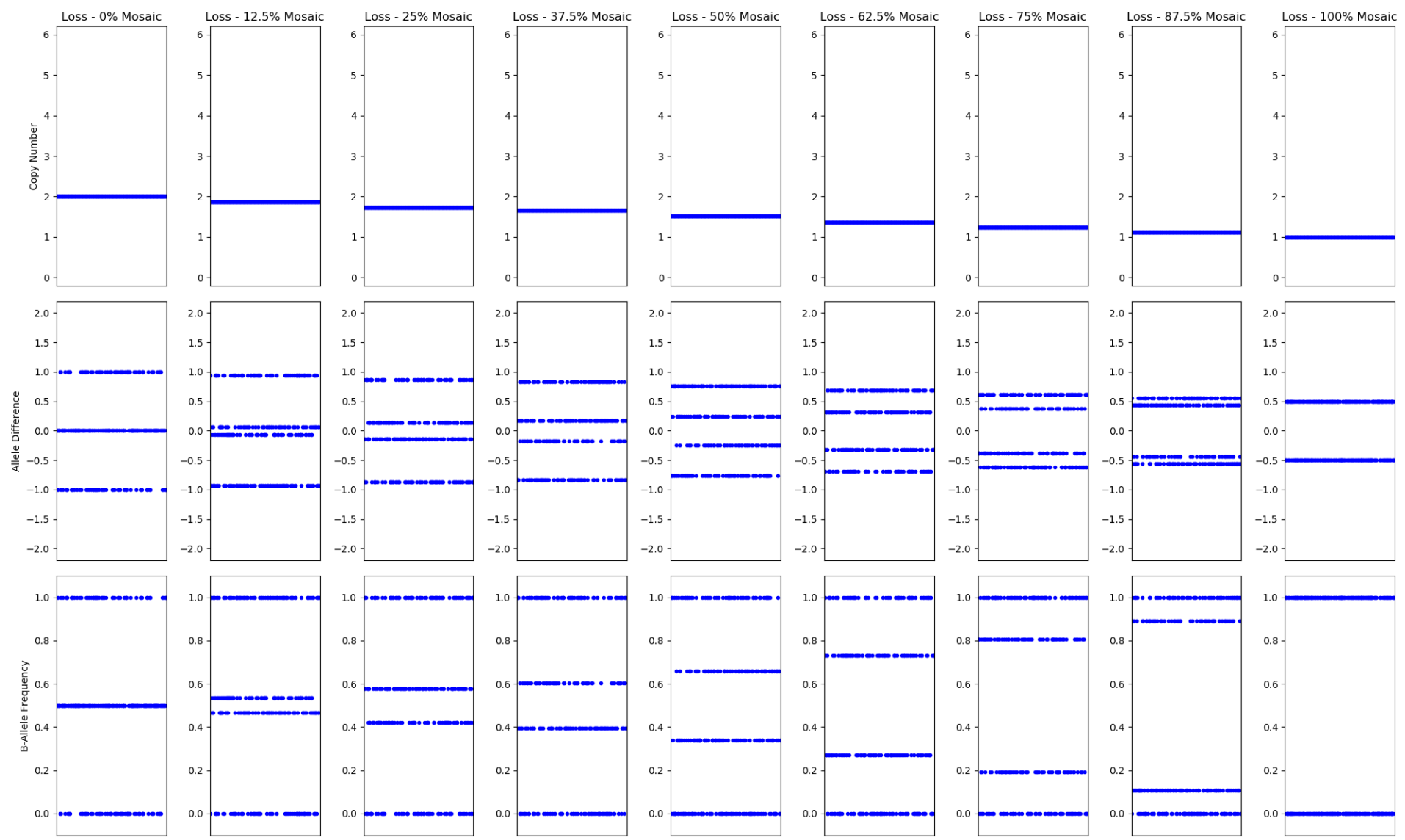
While this model could be used to simulate SNP array data from a large number of cases, 7 aberrations and the mosaic combinations thereof representing the most frequently occurring situations were selected for inclusion in this guide. The abnormalities included are described in detail in Table 1 below.

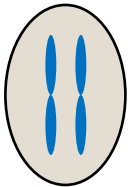
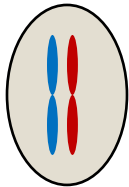
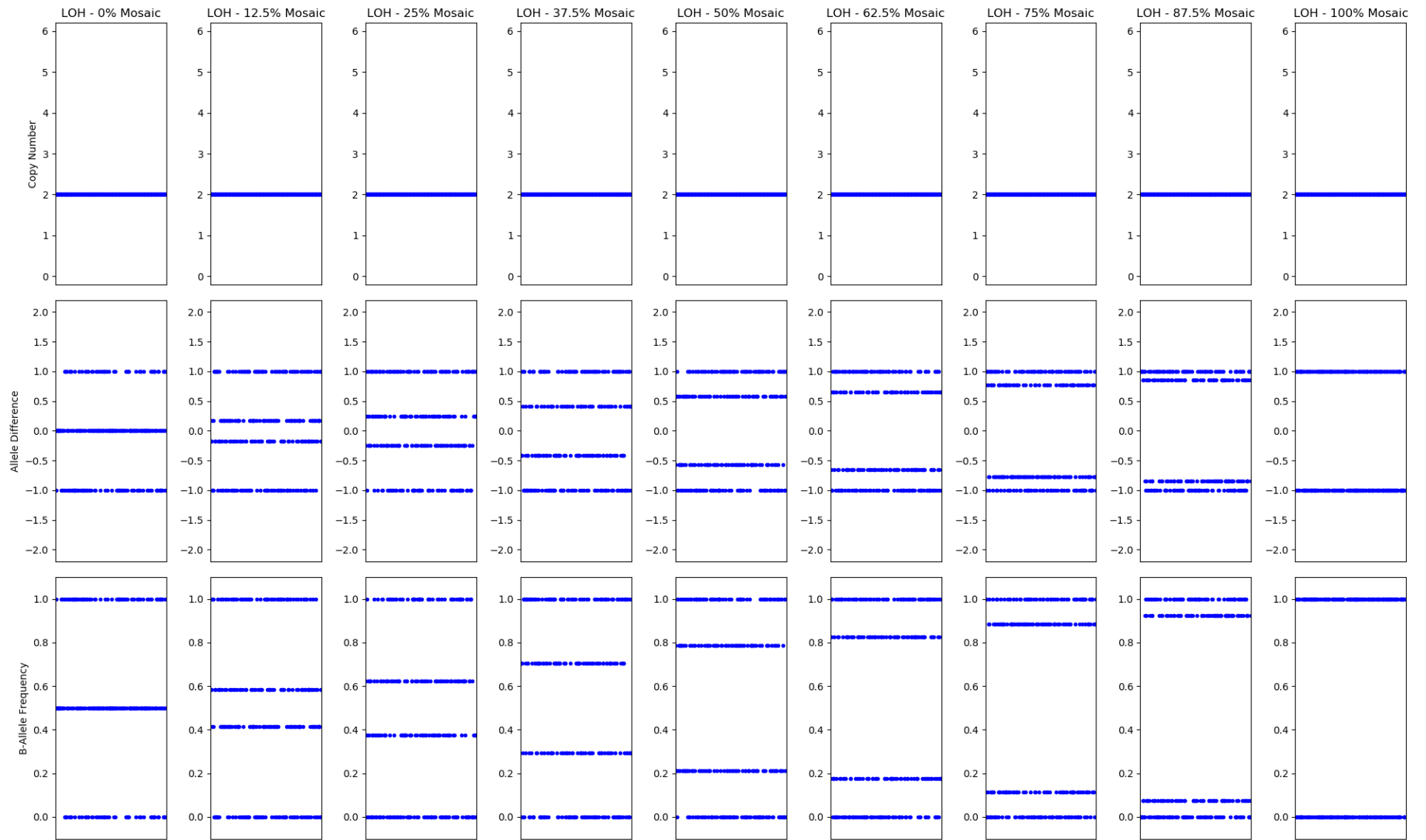
While this guide is a useful companion for the analysis of SNP array data, care should be taken when applying the results of this model as it represents the ideal case. Real-world data may not perfectly match the model output due to system noise, fluorescent dynamic range, variations in the input DNA, etc. Most microarray platforms cannot reliably detect mosaicism below 20%. This guide is not a substitute for proper training and experience in analysis with the specific microarray technology you may be using.

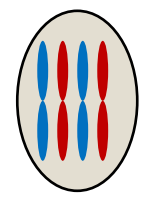
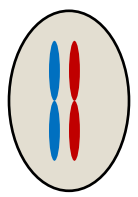
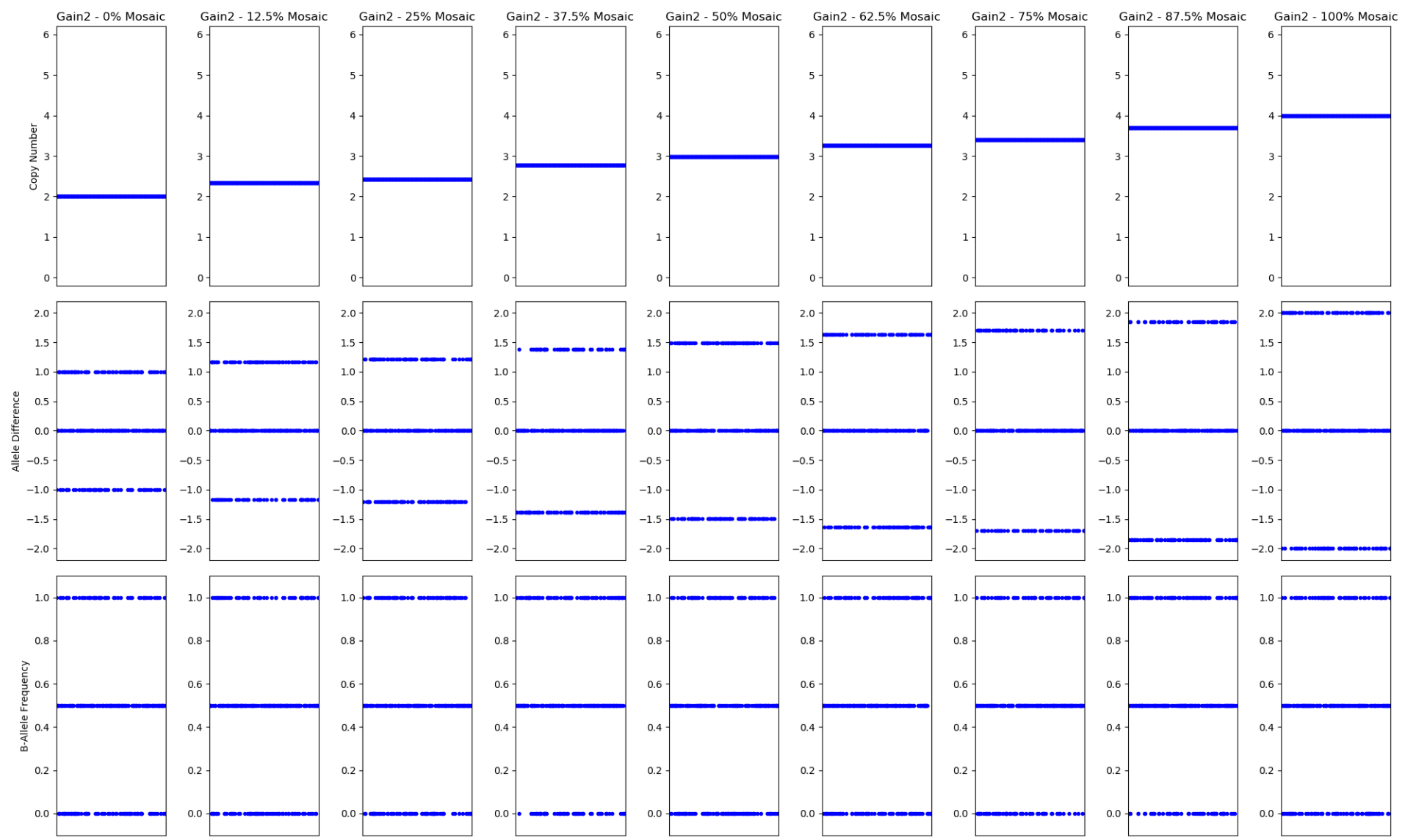
Table 1: The types of cell populations included in this interpretation guide.

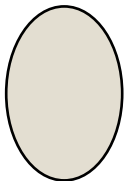
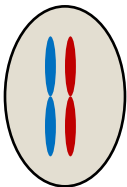
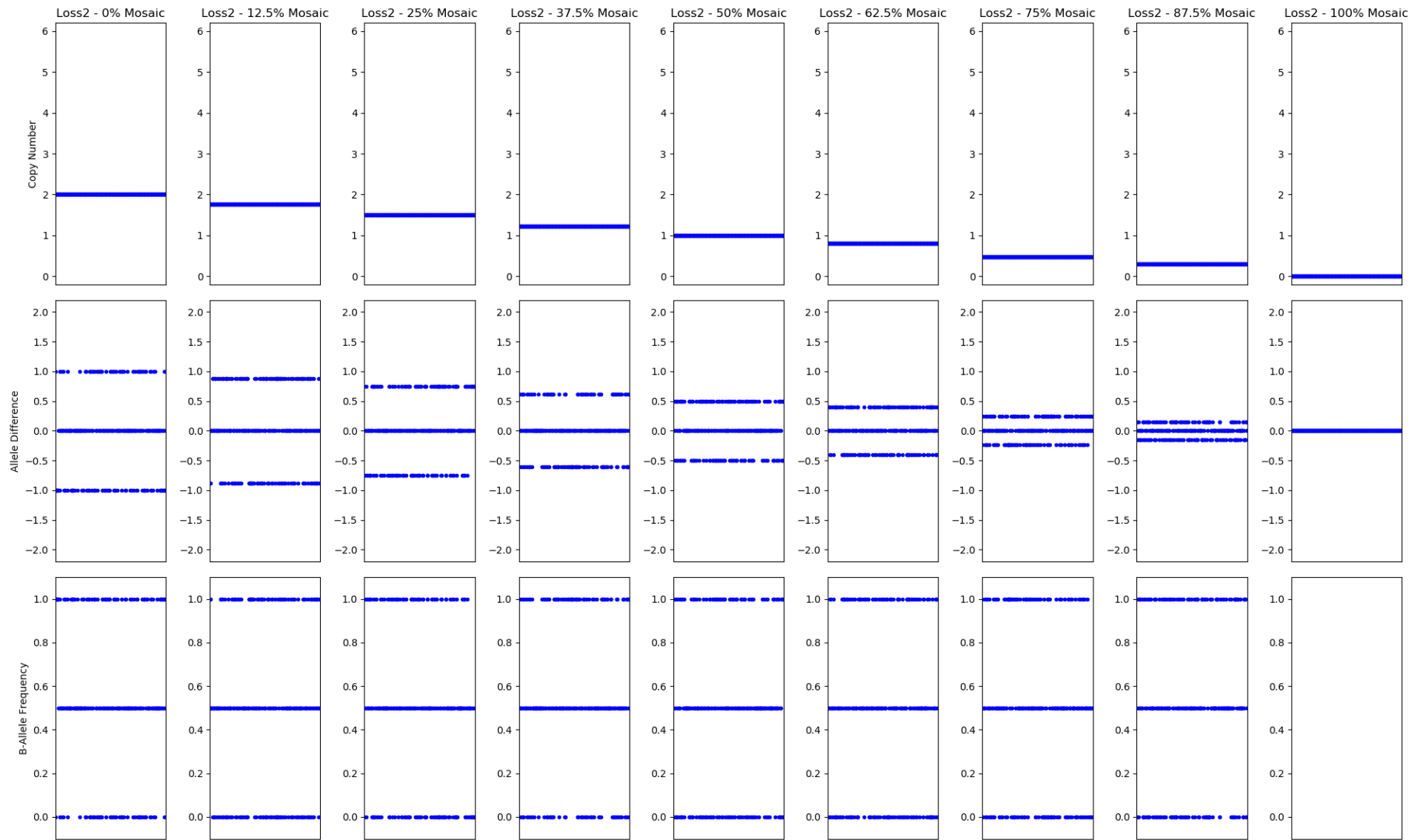
Cell Population	Pictoral Representation	Description
Normal Diploid		This represents the normal two-copy state of most cells. Each cell contains one allele from each parent.
One Copy Loss		This represents a loss of a single copy, leaving only one copy of an allele in the cell.
One Copy Gain		This represents a gain of a single copy, resulting in a two-haplotype state with three total copies.
Loss of Heterozygosity (LOH)		This represents a state where the copy-number is normal, but only a single haplotype is present. This could be due to uniparental disomy, meiotic or mitotic non-disjunction errors, or in cases with significant inbreeding.
Two Copy Gain		This represents a state where the entire diploid content of a cell is duplicated resulting in a tetrasomy state with four total copies and two haplotypes.
Two Copy Loss		This represents a state where both copies of an allele are lost; as in a homozygous deletion.
Three Haplotype State		This represents a single copy gain with three total copies, but also three unique haplotypes. The additional haplotype most likely originates from a parent.
Parental Cell Contamination		This represents a state where the haplotype of the cells matches one of the parents, rather than the offspring.



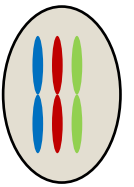
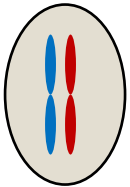
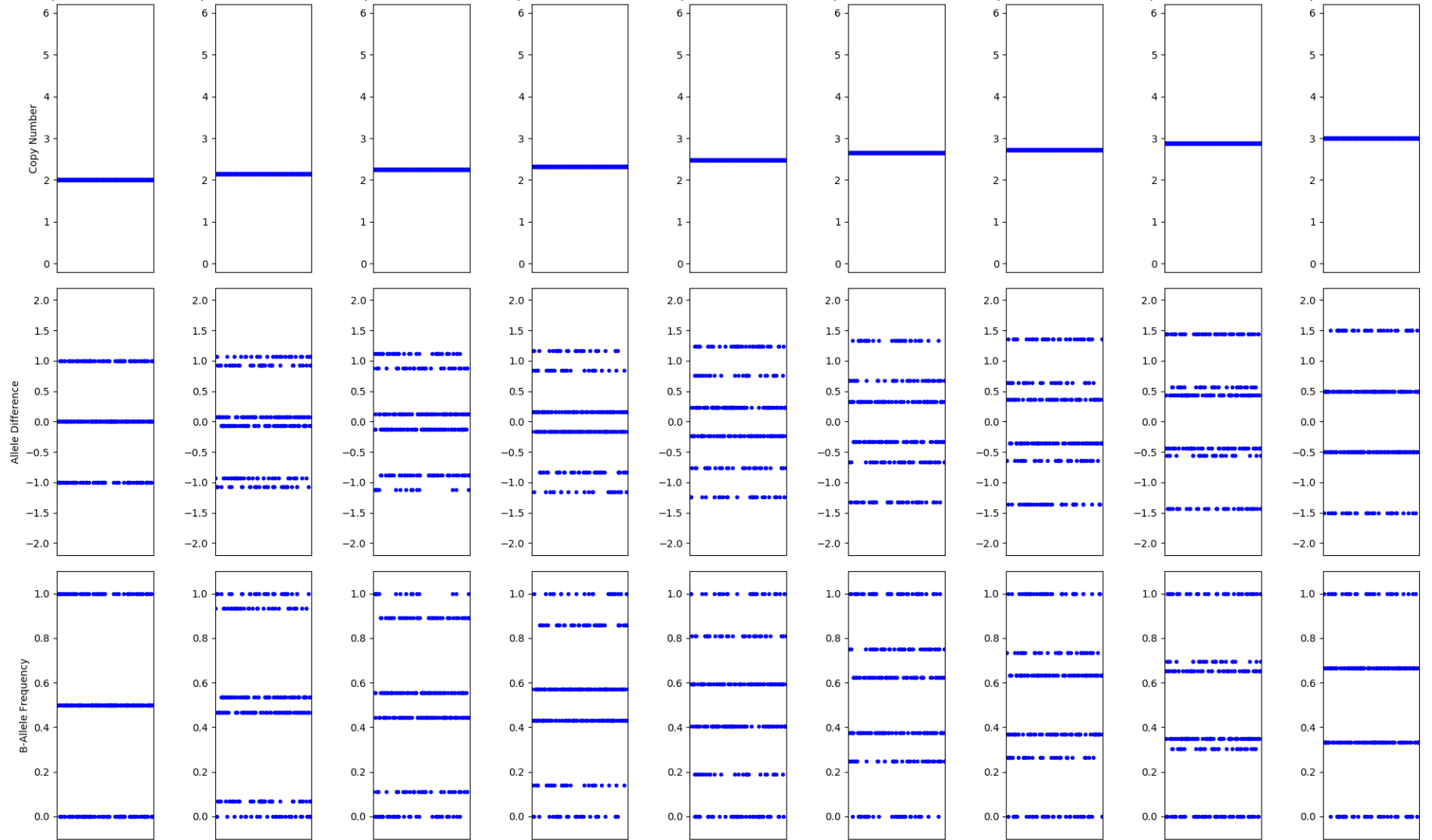








HaploGain - 0% Mosaic HaploGain - 12.5% Mosaic HaploGain - 25% Mosaic HaploGain - 37.5% Mosaic HaploGain - 50% Mosaic HaploGain - 62.5% Mosaic HaploGain - 75% Mosaic HaploGain - 87.5% Mosaic HaploGain - 100% Mosaic



Contamination - 0% Mosaic Contamination - 12.5% Mosaic Contamination - 25% Mosaic Contamination - 37.5% Mosaic Contamination - 50% Mosaic Contamination - 62.5% Mosaic Contamination - 75% Mosaic Contamination - 87.5% Mosaic Contamination - 100% Mosaic

