The file format is composed of a “Table of Contents” at the head of the file. The table of contents has an arbitrary number of tables of content (TOC) entries. This allows for forwards and backwards compatibility as new fields are added to the file format or old fields are removed. A table-of-content entry is 6 bytes long and is made up of two parts: an ID and a file OFFSET. The ID is a two-byte short integer that identifies what the variable is, and the file OFFSET is a four-byte integer that identifies the file offset (relative to the start of the file) where the data for the variable resides. Table 1 describes the format for the file header. Table 2 describes the format for a TOC entry. Table 3 describes the IDs recognized in the gtc file specification. **Note that all multi-byte variables are stored with lowest order byte first.**

Parameter	File Offset	Type	Description
Identifier	0	char[3]	‘gtc’
File Version	3	byte	File version (5 is the version corresponding to this document)
Num Entries	4	int32	Number of table-of-content entries (M)
Entry 0	8	TOC Entry	Table-of-content entry 0
…			Table-of-content entries
Entry M-1	8 + (M-1) * 6	TOC Entry	Table-of-content entry M-1

Table 1: File format header

Parameter	Type	Description
ID	short	TOC variable ID
OFFSET	int32	Offset into the file (relative to start of file)

Table 2: TOC entry

ID	Name	Type	Description
1	NumSNPs	int32	Number of SNPs
2	Ploidy	int32	Ploidy of species
3	Ploidy Type	int32	<table><tr><th>Value</th><th>Description</th></tr><tr><td>1</td><td>Diploid</td></tr><tr><td>2</td><td>Autopolyploid</td></tr><tr><td>3</td><td>Allopolyploid</td></tr></table>
10	Sample Name	string	The sample name
11	Sample Plate	string	The sample plate
12	Sample Well	string	The sample well
100	Cluster File	string	The name of the cluster file
101	SNP Manifest	string	The name of the SNP manifest
200	Imaging Date	string	The imaging date
201	AutoCall Date	string	The AutoCall processing date
300	AutoCall Version	string	The AutoCall version
400	Normalization Transformations	XForm[]	The array of normalization transformation
500	Raw Control X	ushort[]	Raw control green intensities. The length of the vector will be equal to the (sections/sample) * (number of controls). Each control intensity will then be repeated for each section per sample.
501	Raw Control Y	ushort[]	Raw control red intensities. The length of the vector will be equal to the (sections/sample) * (number of controls). Each control intensity will then be repeated for each section per sample.
1000	Raw X intensity values	ushort[]	The array of raw green intensities for every SNP

1001	Raw Y intensity values	ushort[]	The array of raw red intensities for every SNP
1002	Genotypes	byte[]	The array of genotypes (see table 4: genotype mapping below) for every SNP
1003	BaseCalls	BaseCall[]	The array of BaseCalls with respect to the TOP strand for every SNP
1004	Genotype Scores	float[]	The array of GenCall scores for every SNP
1005	*Scanner Data*		*Information about the scanner*
	Scanner Name	string	The name of the scanner
	Pmt Green	int32	
	Pmt Red	int32	
	Scanner Version	string	Version of the scanner software used
	Imaging User	string	Name of the scanner user
1006	Call Rate	float	Calculated call rate of the sample
1007	Estimated Gender	char	'M'-Male, 'F'-Female, 'U'-Unknown
1008	LogR Dev	float	
1009	p10GC	float	The 10th percentile of genotype call scores for this sample. In the calculation of this metrics, scores from intensity only loci or scores beneath the genotype call threshold are **not** considered.
1010	DX	int32	DX flag
1011	*Extended sample data*		
	P50GC	float	The 50th percentile of genotype call scores for this sample. In the calculation of this metrics, scores from intensity only loci or scores beneath the genotype call threshold are **not** considered.
	NumCalls	int32	Number of valid calls
	NumNoCalls	int32	Number of invalid calls
	Num Intensity Only	int32	Number of loci that are "Intensity Only" or "Zeroed"
1012	B Allele Frequencies	float[]	B allele frequencies across loci
1013	LogR Ratios	float[]	LogR ratios across loci
1014	*Intensity percentiles (X)*		
	P05 X	ushort	
	P50 X	ushort	
	P95 X	ushort	
1015	*Intensity percentiles (Y)*		
	P05 Y	ushort	
	P50 Y	ushort	
	P95 Y	ushort	
1016	Sentrix ID	string	The Sentrix barcode.

Table 3: Description of TOC variable IDs

Genotype	Byte value
NC	0
AA	1
AB	2
BB	3

NULL	4
A	5
B	6
AAA	7
AAB	8
ABB	9
BBB	10
AAAA	11
AAAB	12
AABB	13
ABBB	14
BBBB	15
...	...
BBBBBBBB	45

Table 4: Genotype mapping table

If the ID in the TOC entry corresponds to an int variable type (NumSNPs or Ploidy, in our case), then the OFFSET in the TOC entry is the actual value of the unsigned integer and not a file offset. If the ID in the TOC entry corresponds to a string variable type, then the first byte at the file location specified by OFFSET is the length of the string (L) and the L bytes of the string follow as single byte characters. If the ID in the TOC corresponds to an array, then the first four bytes at the location specified by OFFSET are an integer corresponding to the length of the array (N). Each element in the array follows. Arrays of type ushort have N ushorts. Arrays of type byte have N bytes. Arrays of type float have N floats. The other two types require a bit of explanation.

BaseCall

A base call is a multiple-character word. The number of characters in the basecall is always two, regardless of the ploidy. The characters are ‘A’, ‘C’, ‘G’, ‘T’, or ‘-‘ for a no-call. For a diploid genotype, the BaseCall will be the nucleotide genotype (relative to the top strand). For a polyploid genotype, the BaseCall will still be exactly two characters. In this case, the software will report the nucleotide genotype for the A and B alleles on the top strand (in that order). In combination with the AB genotypes, the client will be able to reconstruct the full nucleotide genotype on the top strand.

XForm

An XForm is a 52 byte structure made up of 1 4-byte integer followed by 12 4-byte floating point numbers, summarized in Table 5.

Offset	Column Name	Type
0	version	int
4	offset_x	float
8	offset_y	float
12	scale_x	float
16	scale_y	float
20	shear	float
24	theta	float
28	reserved	float
32	reserved	float

36	reserved	float
40	reserved	float
44	reserved	float
48	reserved	float

Table 5: Normalization Transformation structure

To go from raw coordinates (xraw, yraw) to normalized coordinates (xn, yn), perform the following operations:

tempx = xraw - offset_x

tempy = yraw - offset_y

tempx2 = cos(theta) * tempx + sin(theta) * tempy

tempy2 = -sin(theta) * tempx + cos(theta) * tempy

tempx3 = tempx2 - shear * tempy2

tempy3 = tempy2

xn = tempx3 / scale_x

yn = tempy3 / scale_y

The reason that there is an array of normalization transformations is that a given normalization transformation applies to a subset of the SNPs in a product. You can use the NormID parameter from the map file to identify which normalization transformation to apply to every SNP.