Cowpea - COMPLEX Mutations Analysis

Original Data Extracted from VCF File

In [1]:
```
from VCFtoTable import *
from GTtable import *
from GTplots import *
from GTplot import *
from BarPlots import *
from CTbarPlots import *
from variant_hist import*
from stats import *
from FilterVCF import *
from GTfilter import*
from CTguide import *
```

In [2]:
```
vcf_cowpea = '/home/anibal/genome_files/freebayes~bwa~GCF_004118075.1_ASM411807v
```

In [3]:
```
samples_all, vcf_df, chrom_len = VCFtoTable(vcf_cowpea)
```

In [4]:
```
samples_all
```

Out[4]:
```
array(['CBC1_P1', 'CBC5_A1'], dtype=object)
```

In [5]:
```
progenitor = 'CBC1_P1'
mutant = 'CBC5_A1'
samples = [progenitor, mutant]
samples
```

Out[5]:
```
['CBC1_P1', 'CBC5_A1']
```

In [6]:
```
vcf_df
```

Out[6]:

	CHROM	POS	REF	
0	NC_018051.1	11786	T	
1	NC_018051.1	11801	TCTTCCT	
2	NC_018051.1	11813	AGCC	
3	NC_018051.1	11825	GGTAGGTAAT	A(
4	NC_018051.1	18327	G	
...	...	...	...	
1968687	NC_040289.1	41659114	T	
1968688	NC_040289.1	41659137	G	
1968689	NC_040289.1	41667130	GTTTCA	

	CHROM	POS	REF
1968690	NC_040289.1	41667148	T
1968691	NC_040289.1	41668013	CAGGGTTTAGGGTTTAGGGTTCAGGGTTTAGGGTTTAGGGTTCAGG...

1968692 rows × 14 columns

In [7]:
```
chrom_len
```

Out[7]:

CHROM	LEN
NC_040279.1	42129361
NC_040280.1	33908088
NC_040281.1	65292630
NC_040282.1	42731077
NC_040283.1	48746289
NC_040284.1	34463471
NC_040285.1	40876636
NC_040286.1	38363498
NC_040287.1	43933251
NC_040288.1	41327797
NC_040289.1	41684185
NC_018051.1	152415

PART 0: Raw

Contingency Table - RAW - All Chromosomes - (No 0/0, 0/1, 1/1 Filtered)

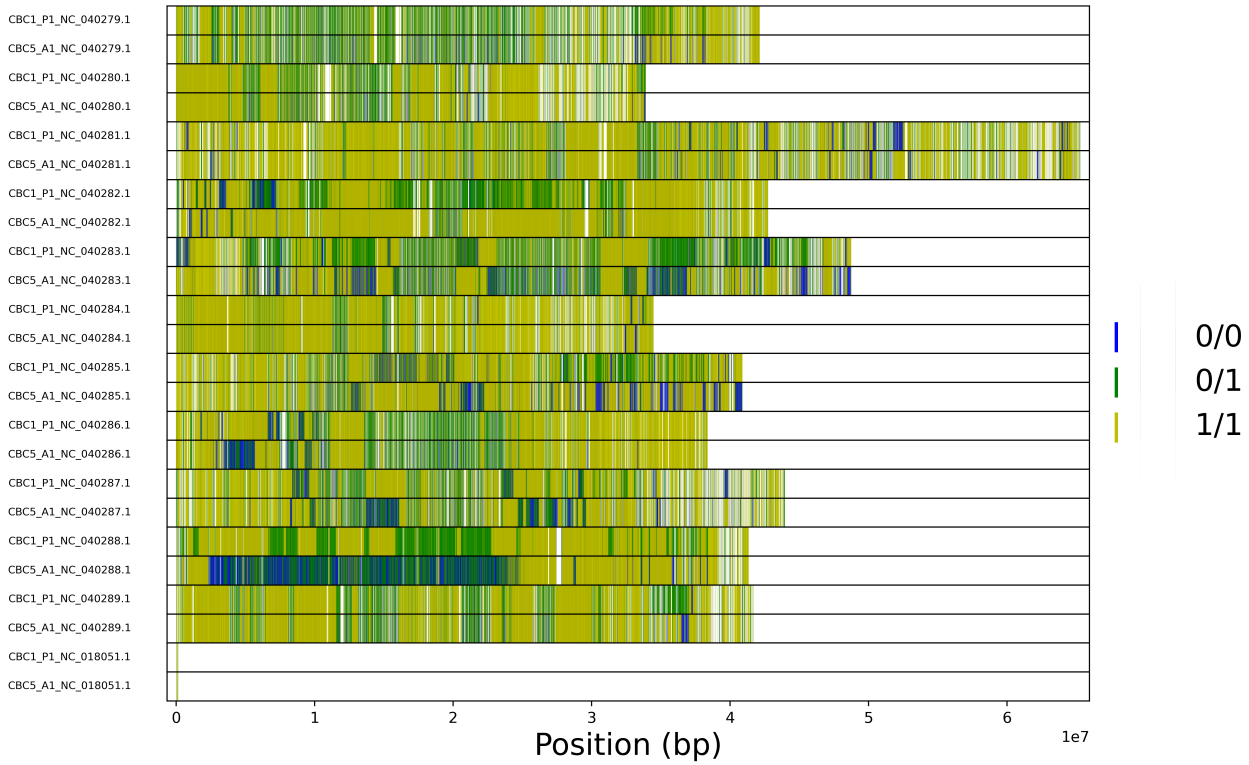
In [8]:
```
contingency_table_0 = contingency_table(samples, vcf_df, 'all')
```

```
Contingency Table - Chromosome all

                CBC5_A1_GT
                      0/0     0/1     1/1  other
CBC1_P1_GT 0/0          0   52496  132507  45048
           0/1     287090  273476  178974  45048
           1/1     211458   19728  767915  45048
           other    45048   45048   45048  45048
```

GT Plot - RAW - All Chromosomes - (No 0/0, 1/1, 'Other' GTs Filtered)

In [9]:
```
plt.close('all')
GTplot(samples, vcf_df, chrom_len)
```

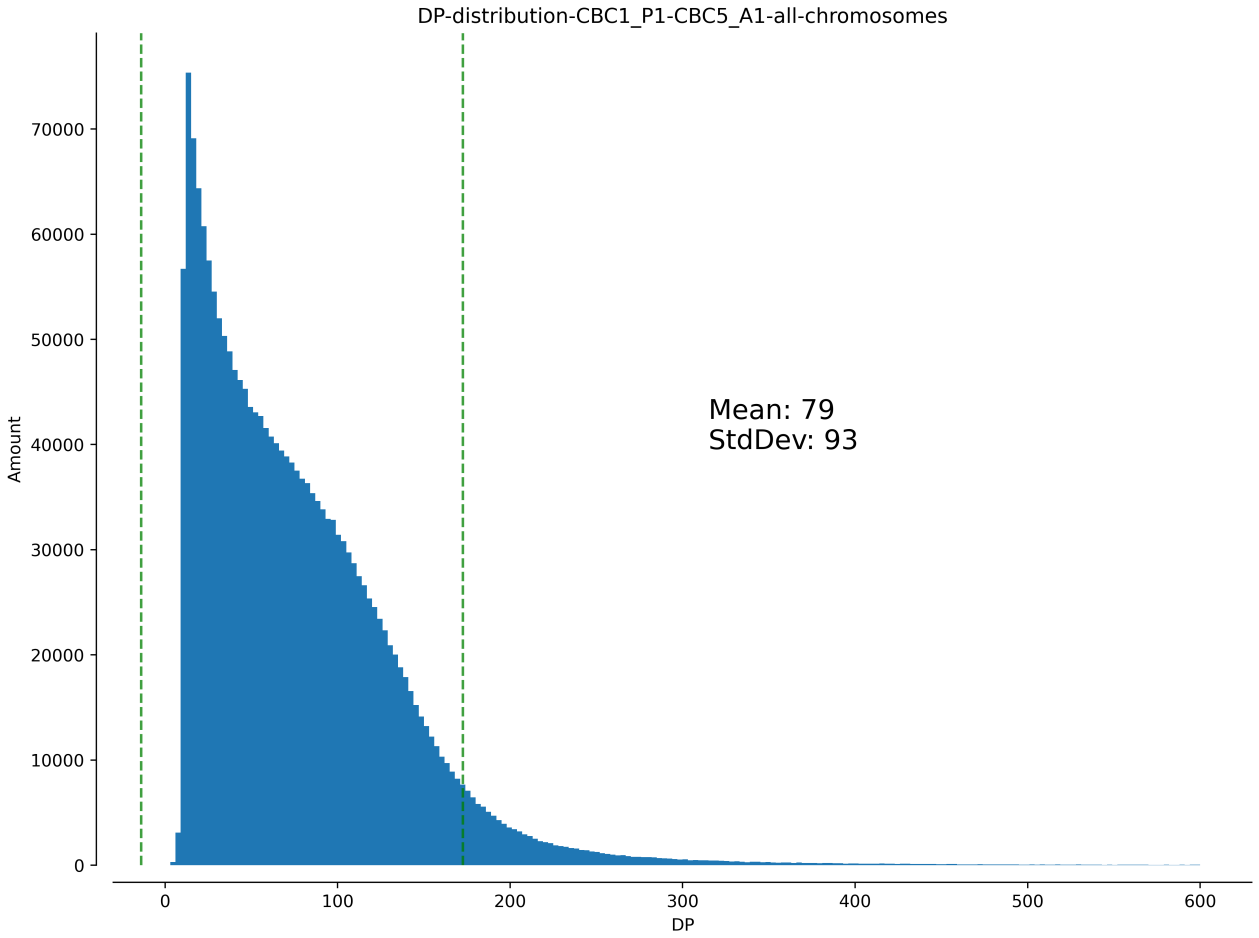
gt-plot-CBC1_P1-CBC5_A1



Histograms - DP , QUAL , TYPE and GT Attributes - All Chromosomes - Unfiltered

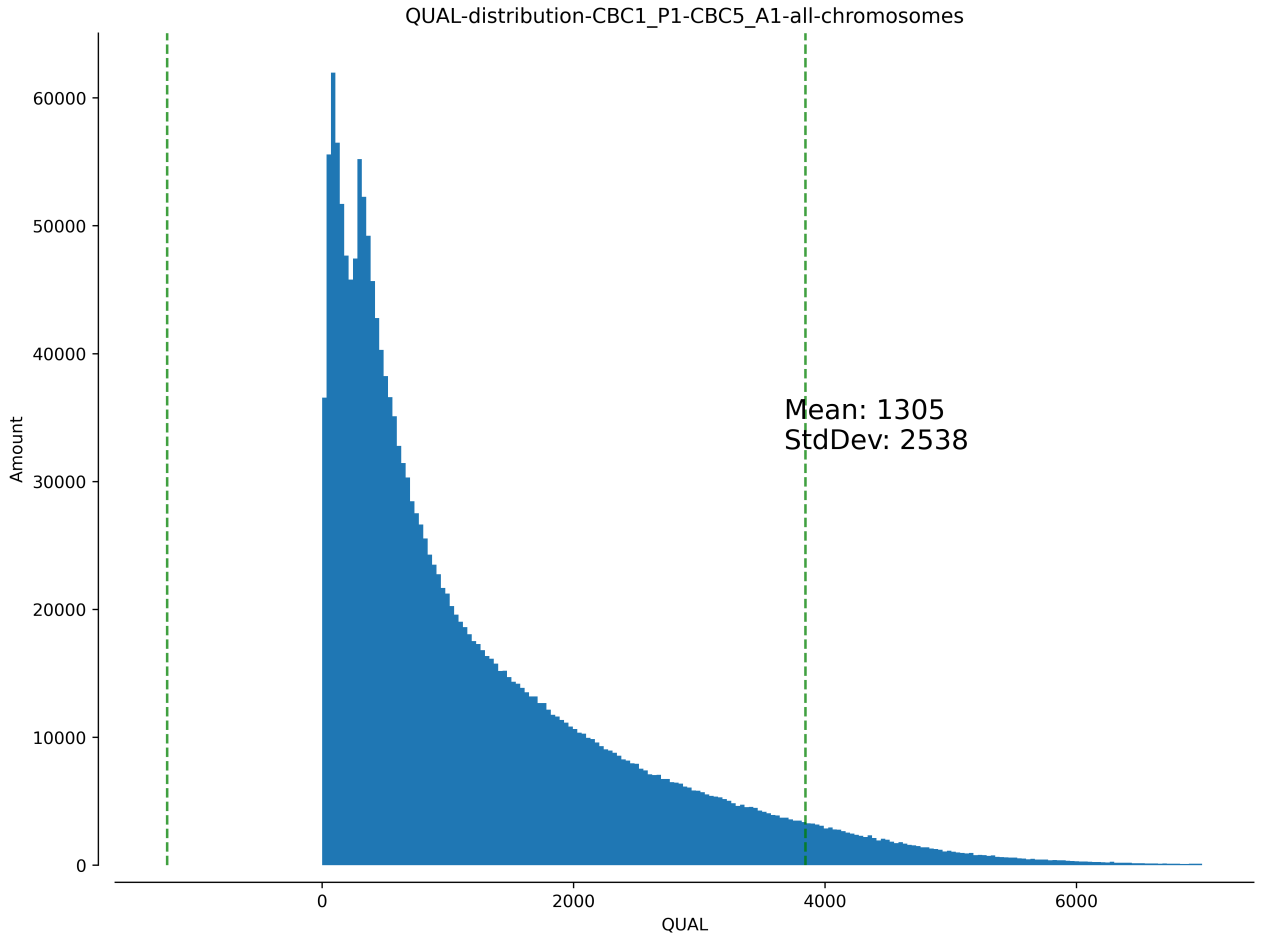
In [10]:

```
plot_variant_hist(samples, vcf_df, 'all', 'DP', bins=200, MSTD=True, xmax=600)
```



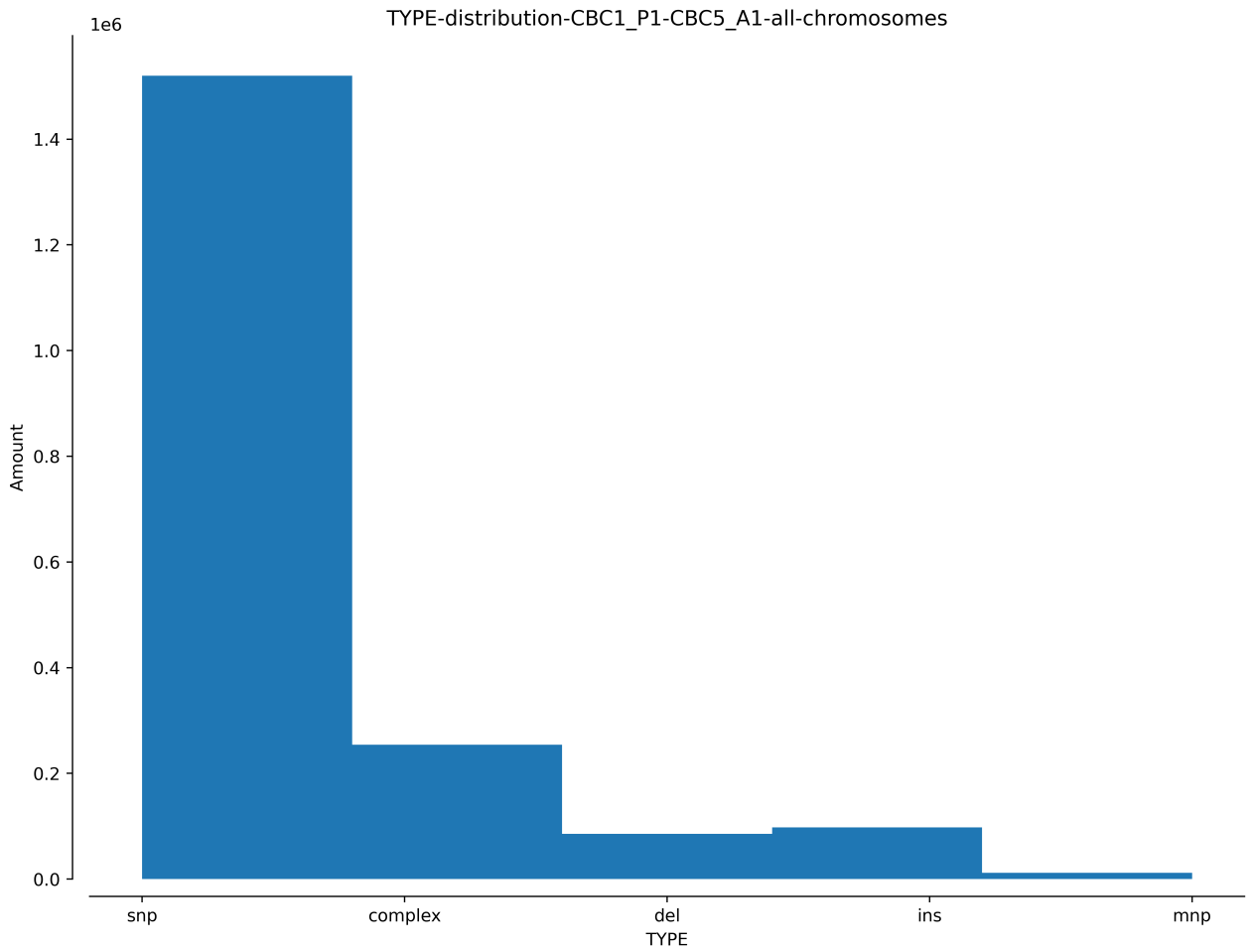
In [11]:

```
plot_variant_hist(samples, vcf_df, 'all', 'QUAL', bins=200,  MSTD=True, xmax=700
```



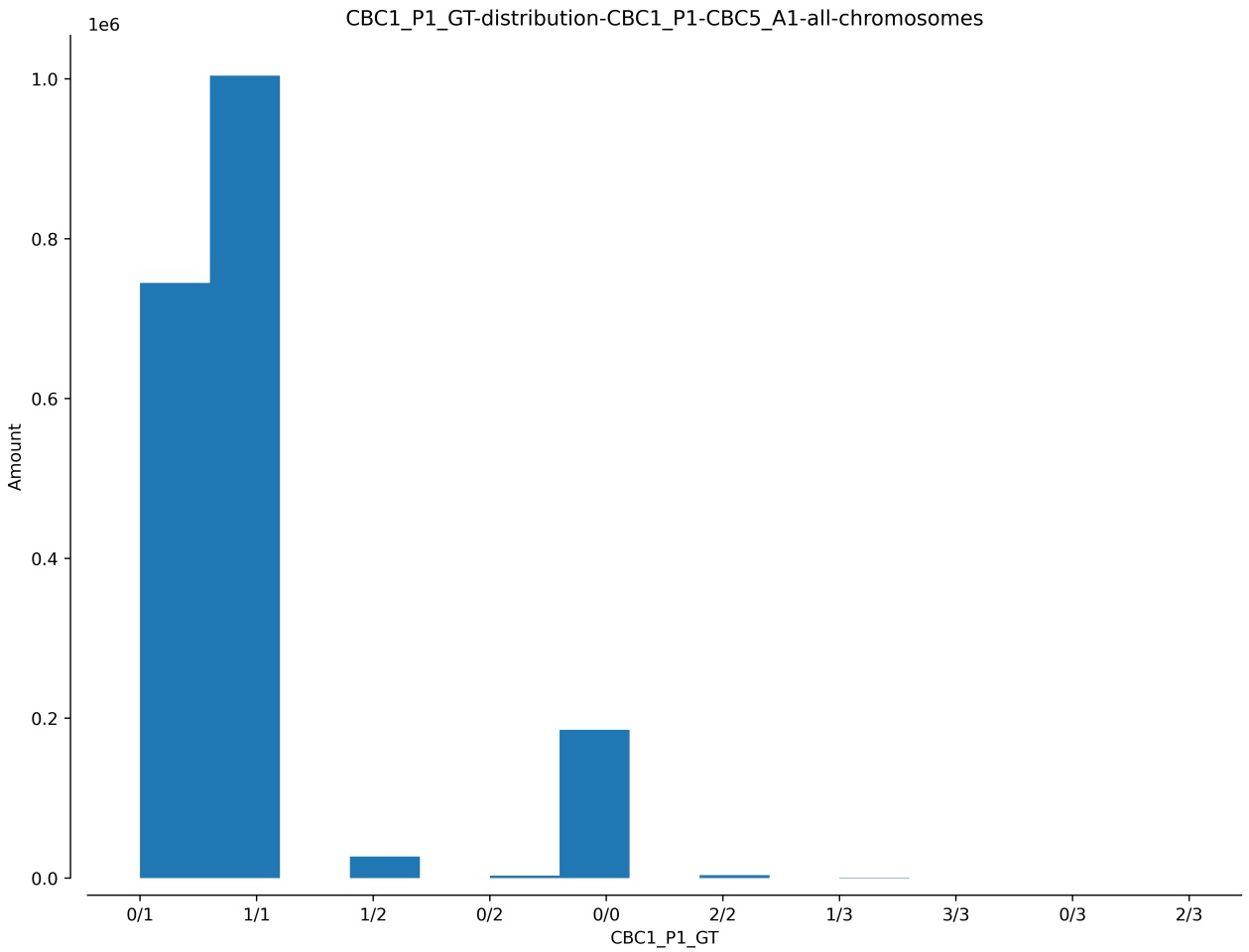
In [12]:

```
plot_variant_hist(samples, vcf_df, 'all', 'TYPE', bins=5)
```



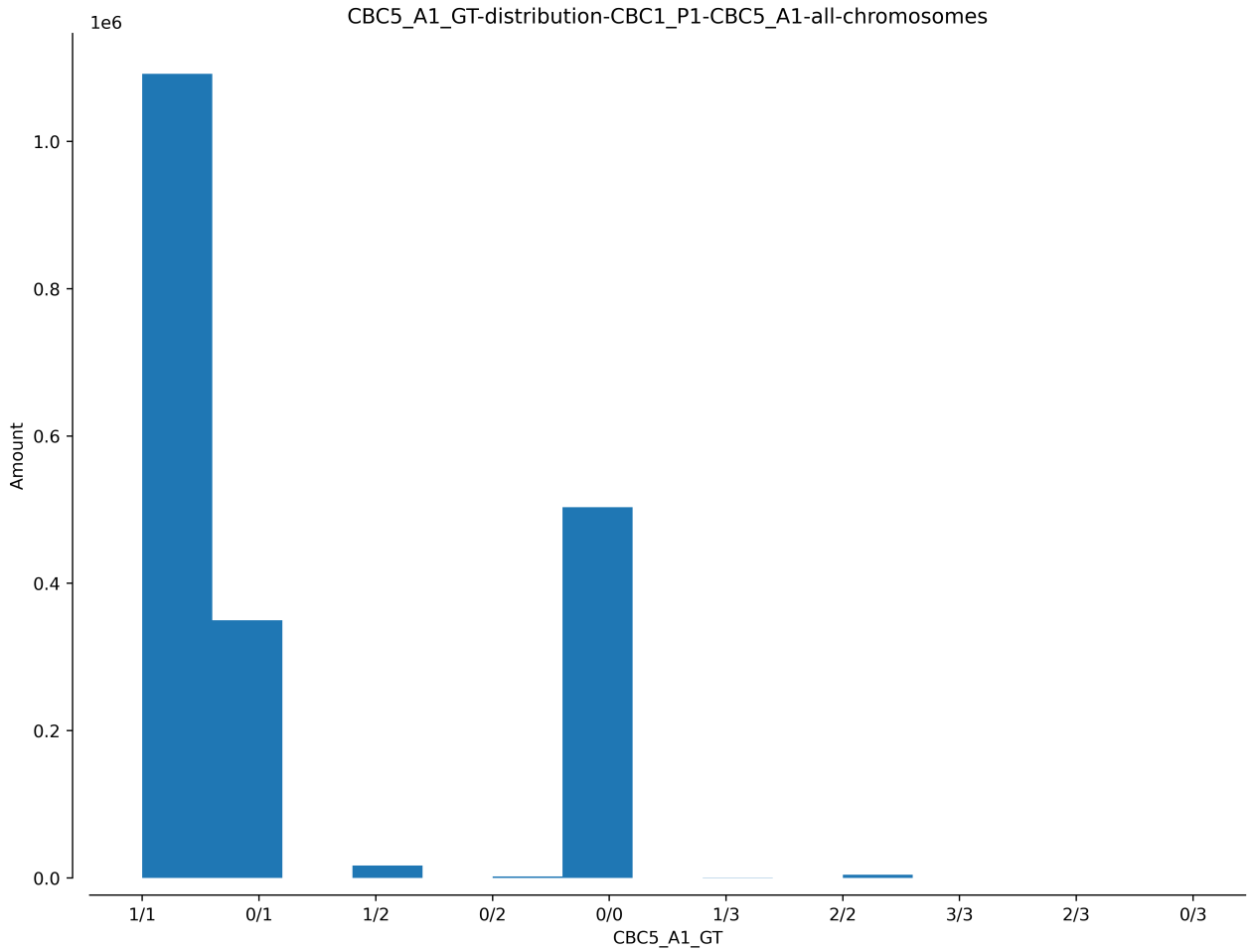
In [15]:

```
plot_variant_hist(samples, vcf_df, 'all', 'CBC1_P1_GT', bins=15)
```



In [16]:

```
plot_variant_hist(samples, vcf_df, 'all', 'CBC5_A1_GT', bins=15)
```



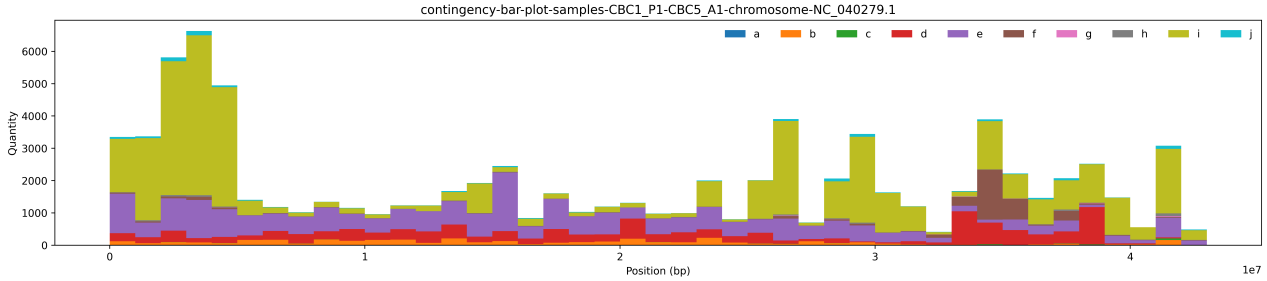
Stacked Bar Plots - RAW

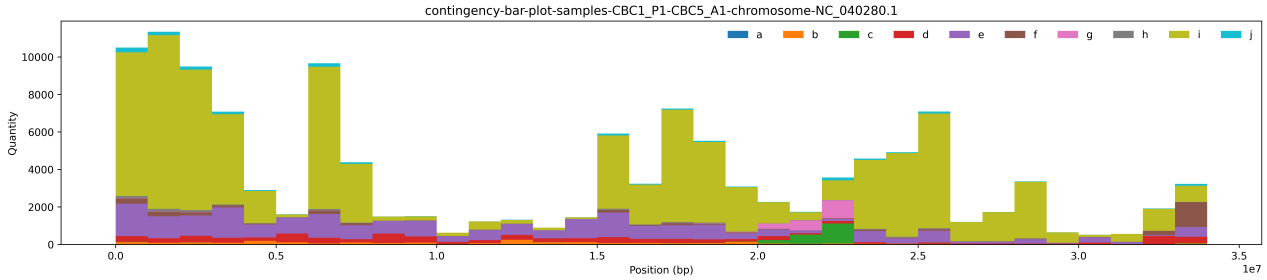
In [17]:
```
ct_guide()
```

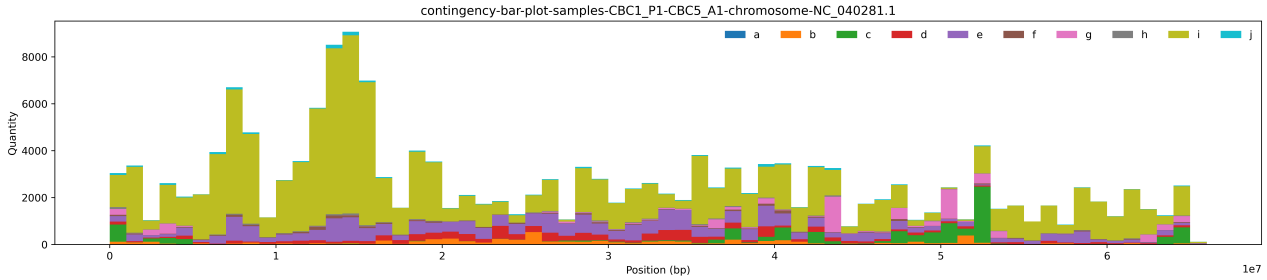
Out[17]:

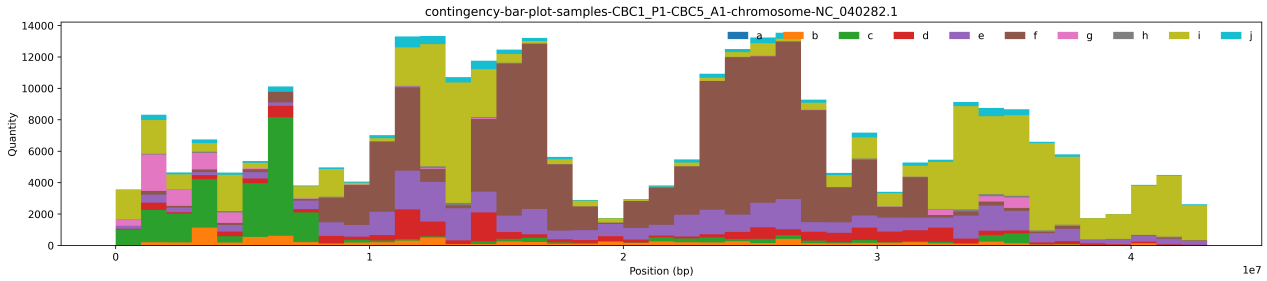
		Mutant			
		0/0	**0/1**	**1/1**	**other**
Progenitor	**0/0**	a	b	c	
	0/1	d	e	f	
	1/1	g	h	i	
	other				j

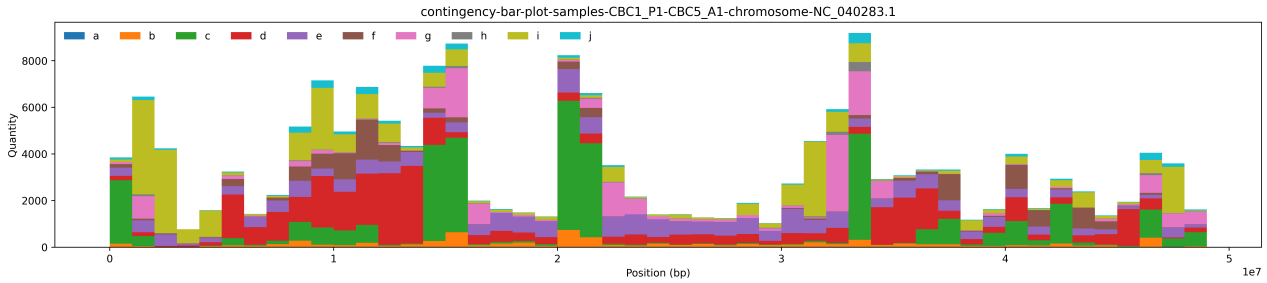
In [18]:
```
plt.close('all')
window_size = 1000000
CTbarPlots(samples, vcf_df, chrom_len, window_size)
```

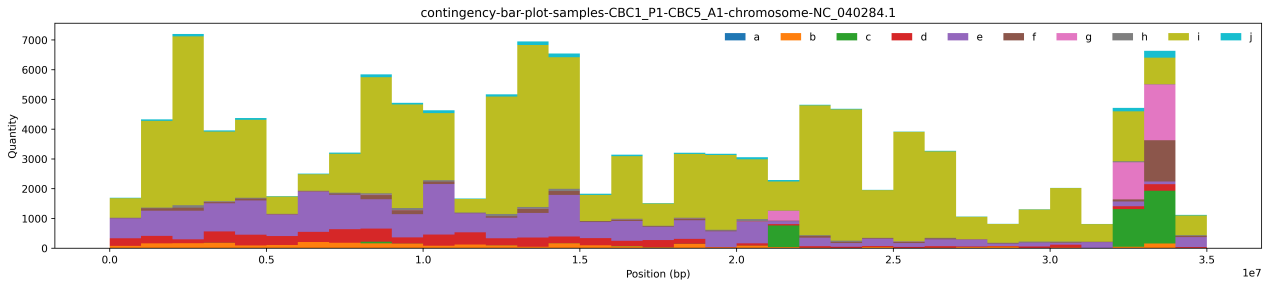

contingency-bar-plot-samples-CBC1_P1-CBC5_A1-chromosome-NC_040279.1
a b c d e f g h i j
Quantity
0 1000 2000 3000 4000 5000 6000
0 1 2 3 4
Position (bp)
1e7

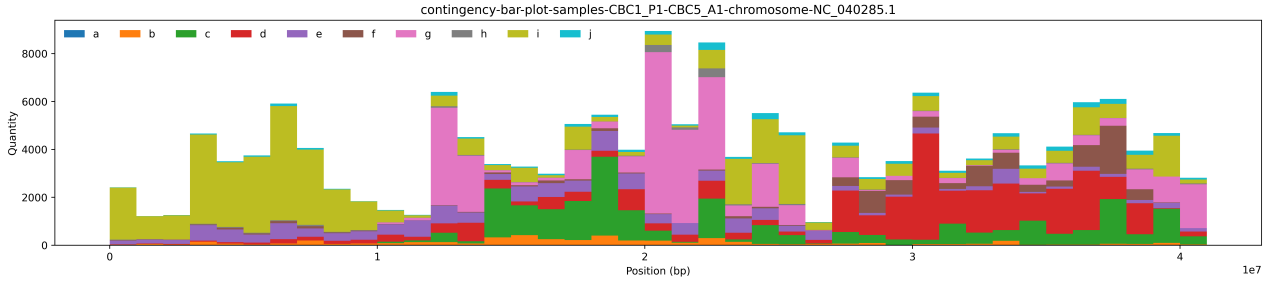

contingency-bar-plot-samples-CBC1_P1-CBC5_A1-chromosome-NC_040280.1
a b c d e f g h i j
Quantity
0 2000 4000 6000 8000 10000
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5
Position (bp)
1e7

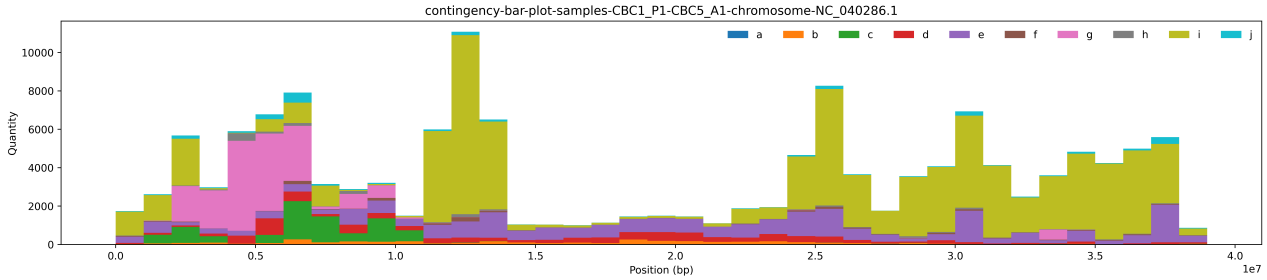

contingency-bar-plot-samples-CBC1_P1-CBC5_A1-chromosome-NC_040281.1
a b c d e f g h i j
Quantity
0 2000 4000 6000 8000
0 1 2 3 4 5 6
Position (bp)
1e7

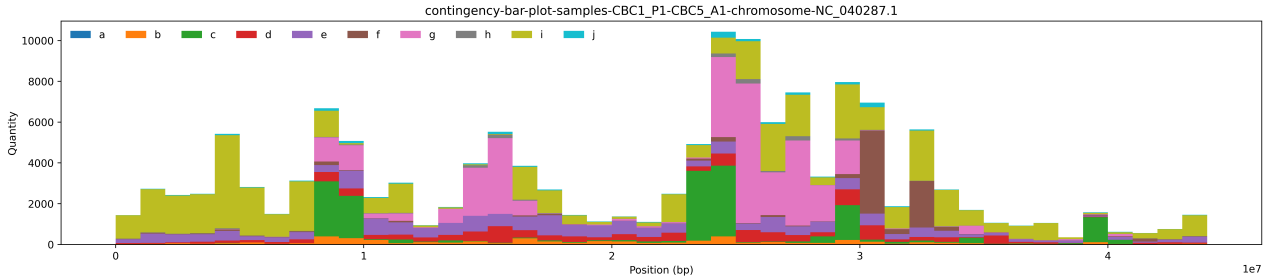

contingency-bar-plot-samples-CBC1_P1-CBC5_A1-chromosome-NC_040282.1
a b c d e f g h i j
Quantity
0 2000 4000 6000 8000 10000 12000 14000
0 1 2 3 4
Position (bp)
1e7

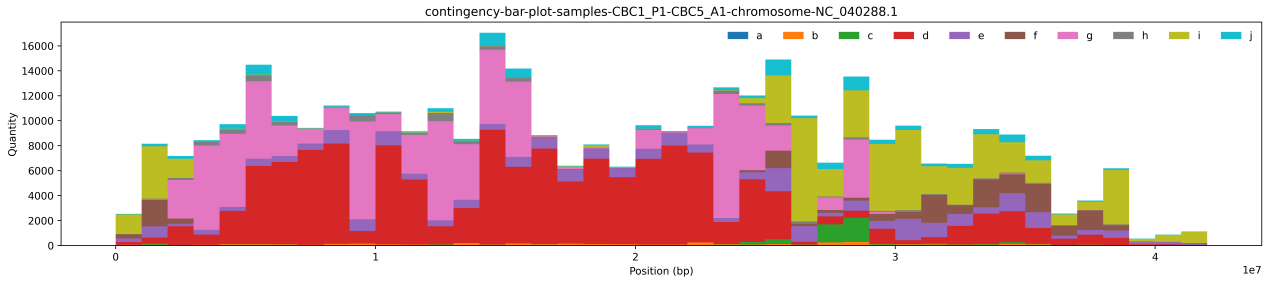

contingency-bar-plot-samples-CBC1_P1-CBC5_A1-chromosome-NC_040283.1
a b c d e f g h i j
Quantity
0 2000 4000 6000 8000
0 1 2 3 4 5
Position (bp)
1e7

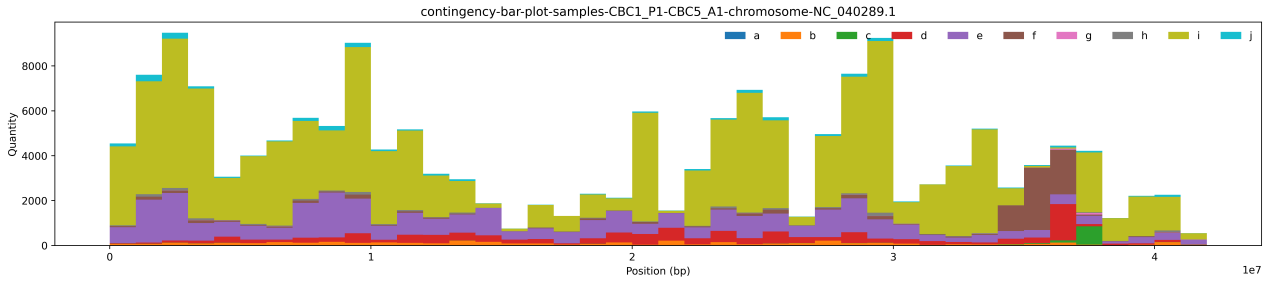

contingency-bar-plot-samples-CBC1_P1-CBC5_A1-chromosome-NC_040284.1
a b c d e f g h i j
Quantity
0 1000 2000 3000 4000 5000 6000 7000
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5
Position (bp)
1e7











PART 1: Filter Out Mitochondria and Chloroplast Chromosomes

Drop Mitochrondria and Chloroplast Chromosomes from `vcf_df` and `chrom_len`

In [19]:

```
drop_mito_chloro = "CHROM != NC_018051.1"
vcf_df_00 = filter_vcf(vcf_df, drop_mito_chloro)
vcf_df_00
```

Out[19]:

	CHROM	POS	REF

	CHROM	POS	REF	
0	NC_040279.1	912	G	
1	NC_040279.1	948	AGGGGAAAC	A(
2	NC_040279.1	1173	C	
3	NC_040279.1	1390	C	
4	NC_040279.1	1424	T	
...	...	...	...	
1968602	NC_040289.1	41659114	T	
1968603	NC_040289.1	41659137	G	
1968604	NC_040289.1	41667130	GTTTCA	
1968605	NC_040289.1	41667148	T	
1968606	NC_040289.1	41668013	CAGGGTTTAGGGTTTAGGGTTCAGGGTTTAGGGTTTAGGGTTCAGG...	

1968607 rows × 14 columns

In [20]:
```
mito_chloro = ['NC_018051.1']
chrom_len_00 = chrom_len.drop(mito_chloro)
chrom_len_00
```

Out[20]:

CHROM	LEN
NC_040279.1	42129361
NC_040280.1	33908088
NC_040281.1	65292630
NC_040282.1	42731077
NC_040283.1	48746289
NC_040284.1	34463471
NC_040285.1	40876636
NC_040286.1	38363498
NC_040287.1	43933251
NC_040288.1	41327797
NC_040289.1	41684185

Create Mitochrondria and Chloroplast Variants and Chomosome Length Dataframes

In [21]:
```
drop_chrom = "CHROM!=NC_040279.1, CHROM!=NC_040280.1, CHROM!=NC_040281.1, CHROM!
vcf_df_mito_chloro = filter_vcf(vcf_df, drop_chrom)
vcf_df_mito_chloro
```

Out[21]:

	CHROM	POS	REF	ALT	QUAL	DP	CBC1_P1_DP	CBC5_A1_DP
0	NC_018051.1	11786	T	C	492.441010	22	9	13
1	NC_018051.1	11801	TCTTCCT	CCTACCC	319.031006	26	10	16
2	NC_018051.1	11813	AGCC	GGCT	305.860992	28	10	18
3	NC_018051.1	11825	GGTAGGTAAT	AGTGGGGAAC	315.924988	28	9	19
4	NC_018051.1	18327	G	A	580.976990	25	15	10
...	...	...	...	...	...	...	...	...
80	NC_018051.1	129428	T	C	245.087997	20	10	10
81	NC_018051.1	132082	C	G	586.130981	33	16	17
82	NC_018051.1	132156	C	T	45.394600	19	10	9
83	NC_018051.1	132169	CCGGT	ACGGG	478.346985	19	9	10
84	NC_018051.1	132184	A	G	21.932899	19	9	10

85 rows × 14 columns

In [22]:

```
mito_chloro_len = chrom_len.loc[mito_chloro]
mito_chloro_len
```

Out[22]:

	LEN
CHROM	
NC_018051.1	152415

Contingency Table - No Mitochondria/Chloroplast

In [23]:

```
contingency_table_1 = contingency_table(samples, vcf_df_00, 'all')
```

```
Contingency Table - Chromosome all

                CBC5_A1_GT
                      0/0     0/1     1/1  other
CBC1_P1_GT 0/0          0   52496  132507  45041
           0/1     287090  273450  178971  45041
           1/1     211458   19719  767875  45041
           other    45041   45041   45041  45041
```

GT Plot - No Mitochondria/Chloroplast

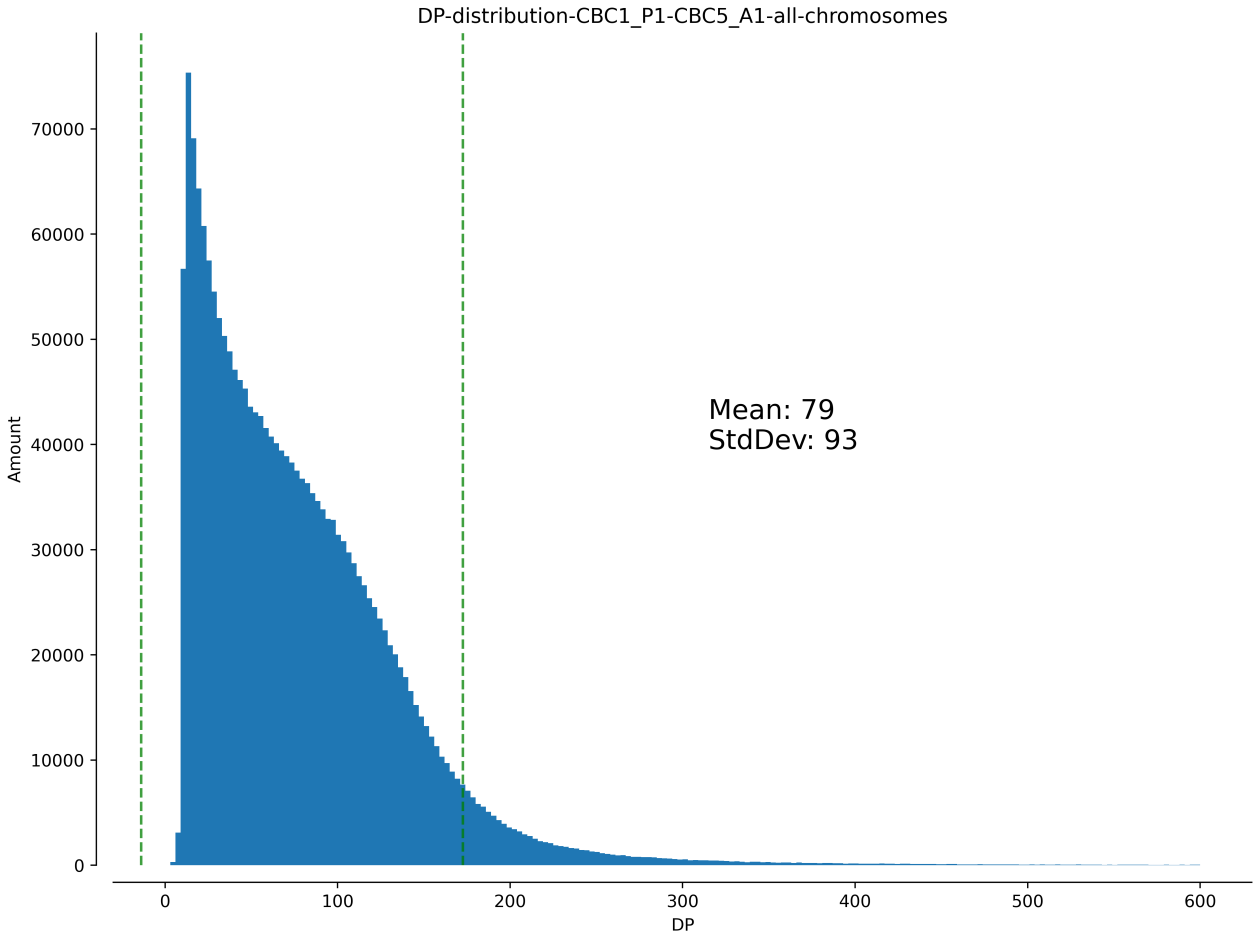
In [24]:

```
# plt.close('all')
# GTplot(samples, vcf_df_00, chrom_len_00)
```

Histograms - DP , QUAL , TYPE and GT Attributes - No Mitochondria/Chloroplast

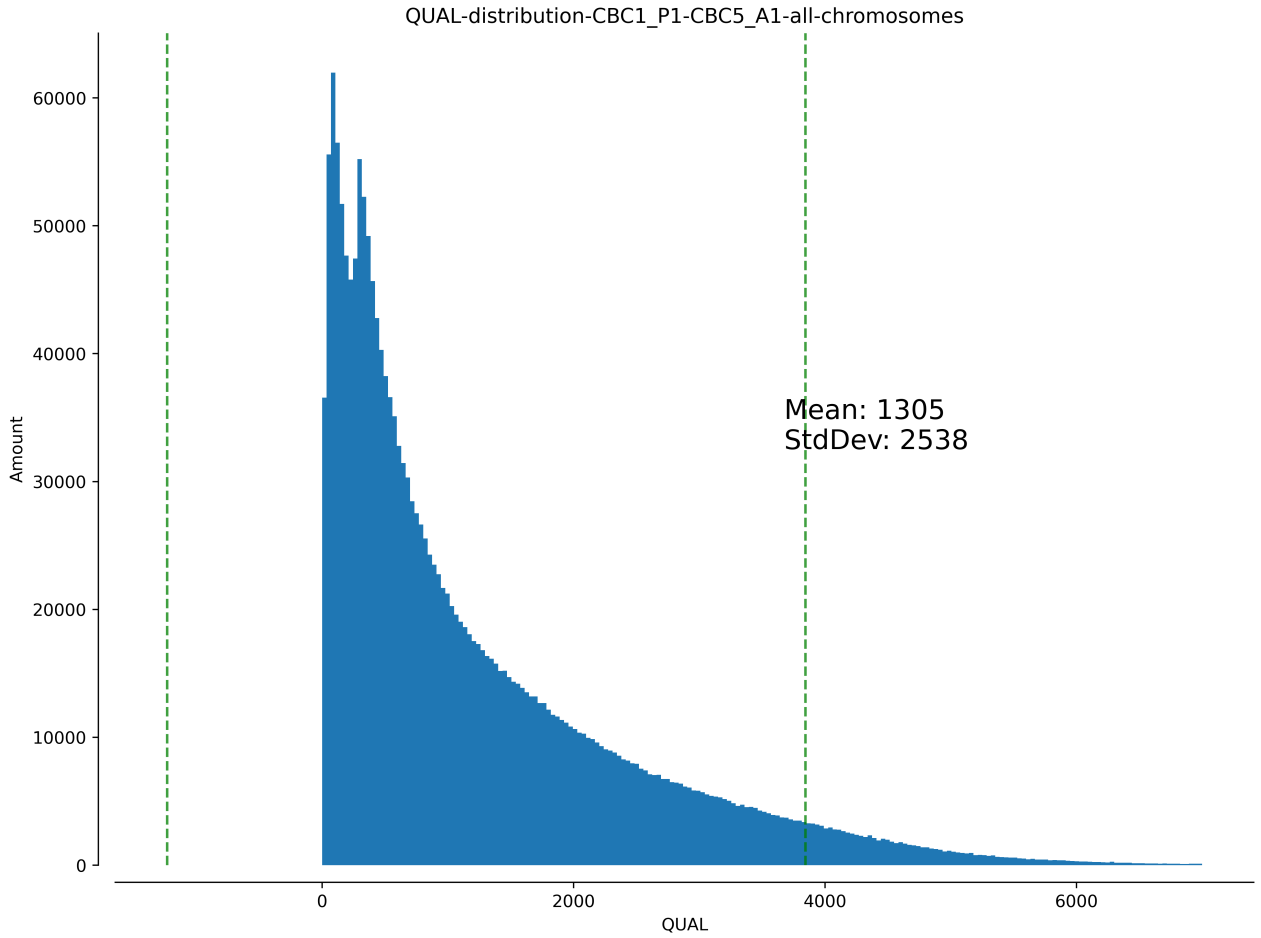
In [25]:

```
plot_variant_hist(samples, vcf_df_00, 'all', 'DP', bins=200, MSTD=True, xmax=600
```



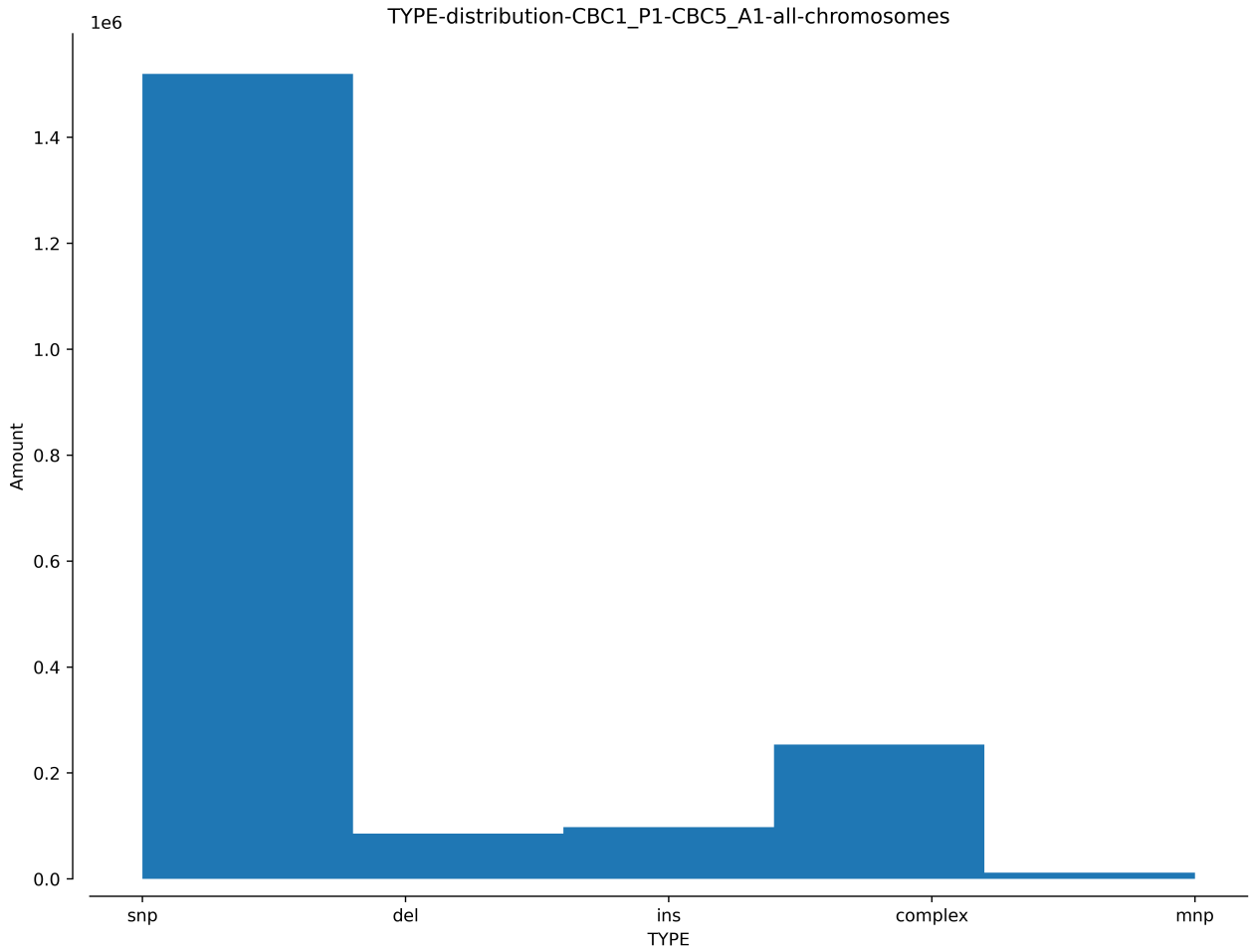
In [26]:

```
plot_variant_hist(samples, vcf_df_00, 'all', 'QUAL', bins=200,  MSTD=True, xmax=
```



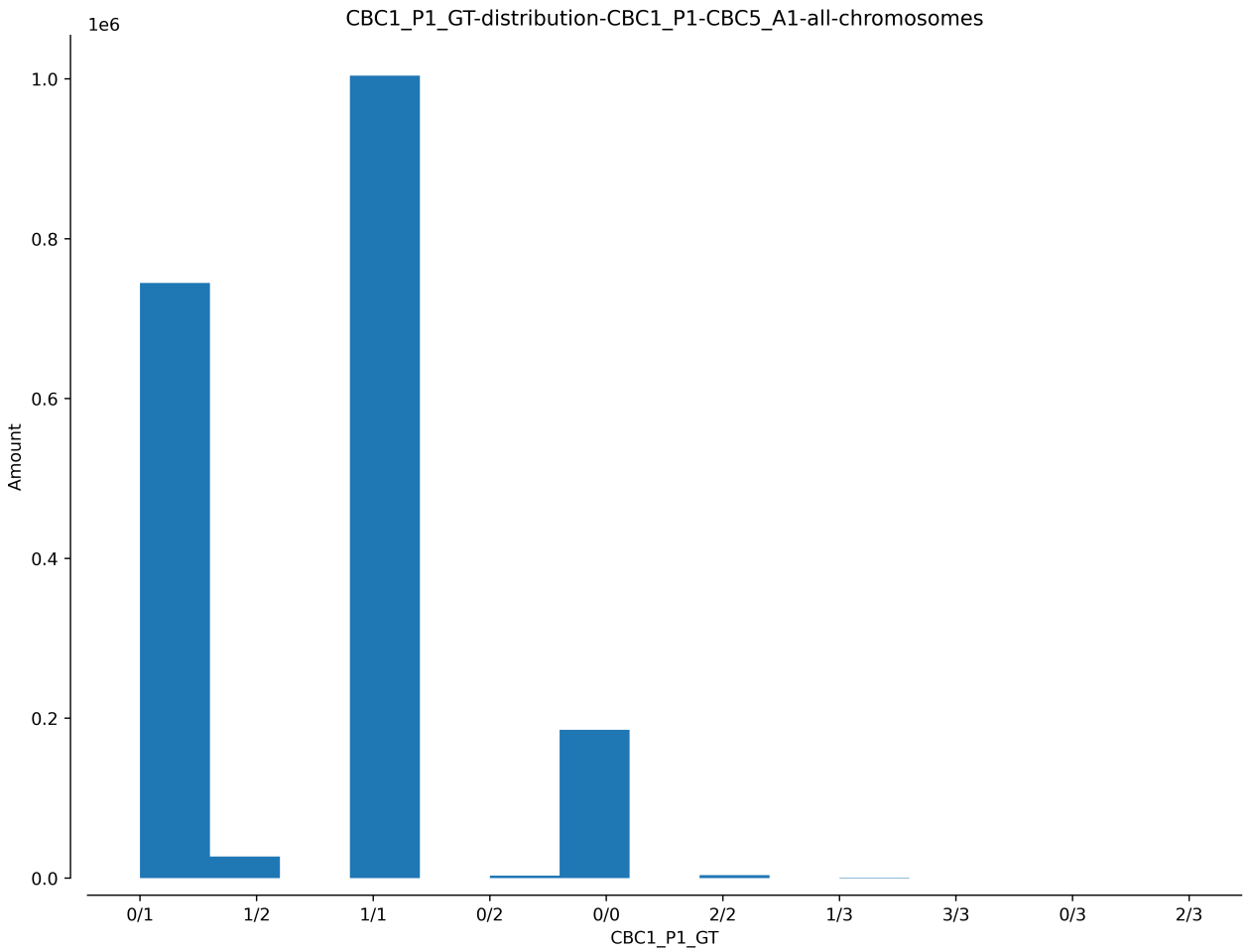
In [27]:

```
plot_variant_hist(samples, vcf_df_00, 'all', 'TYPE', bins=5)
```



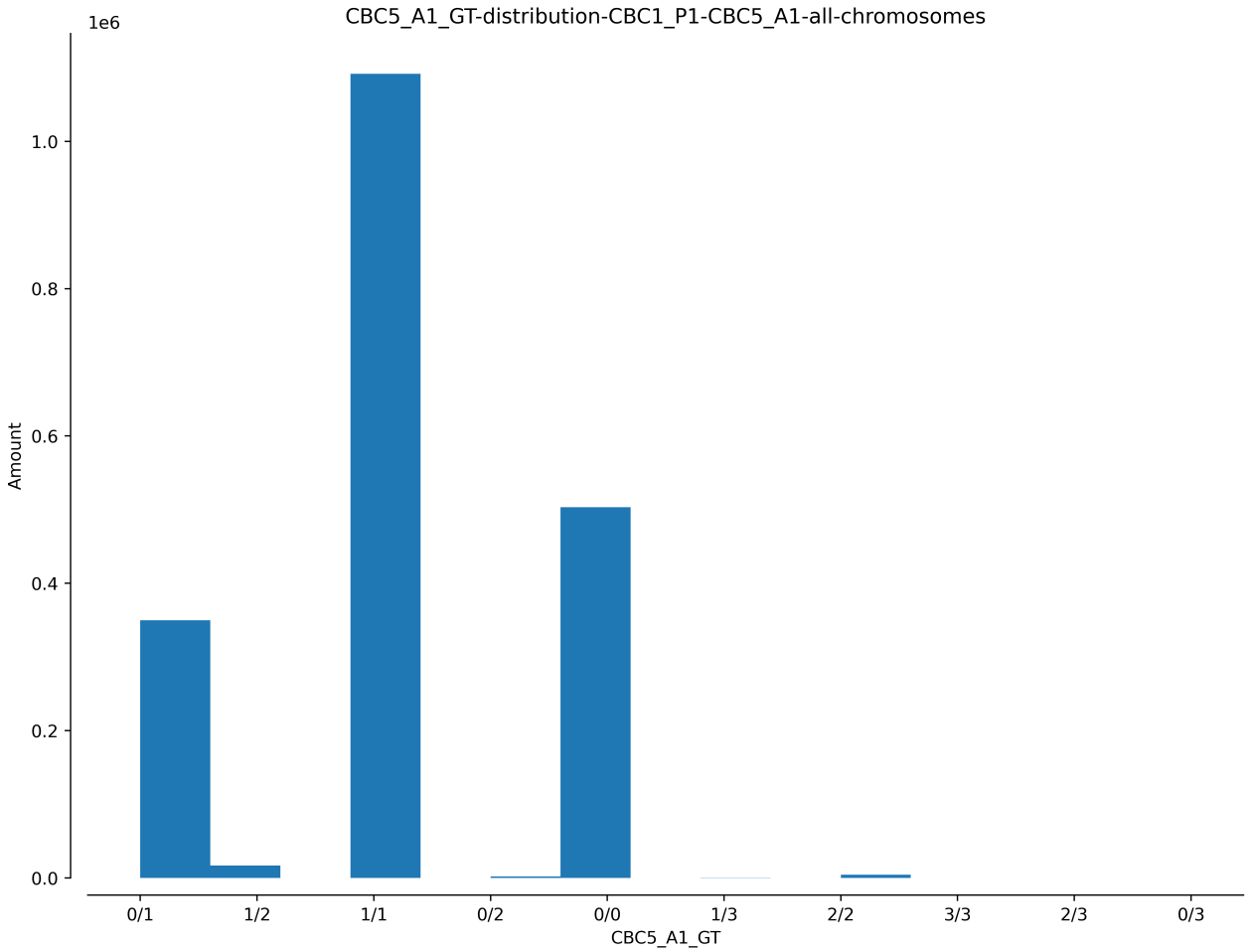
In [29]:

```
plot_variant_hist(samples, vcf_df_00, 'all', '%s_GT' % progenitor, bins=15)
```



In [30]:

```
plot_variant_hist(samples, vcf_df_00, 'all', '%s_GT' % mutant, bins=15)
```



PART 2: Cutting Off by Mean±2StdDev Histograms of DP Attribute

In [31]:
```
cutoff_left = vcf_df_00.DP.mean() - (2 * vcf_df_00.DP.std())
cutoff_right = vcf_df_00.DP.mean() + (2 * vcf_df_00.DP.std())

filter_dp = "DP >= %i, DP <= %i" % (cutoff_left, cutoff_right)
print(filter_dp)

vcf_df_01 = filter_vcf(vcf_df_00, filter_dp)
vcf_df_01
```

```
DP >= -107, DP <= 265
```

Out[31]:

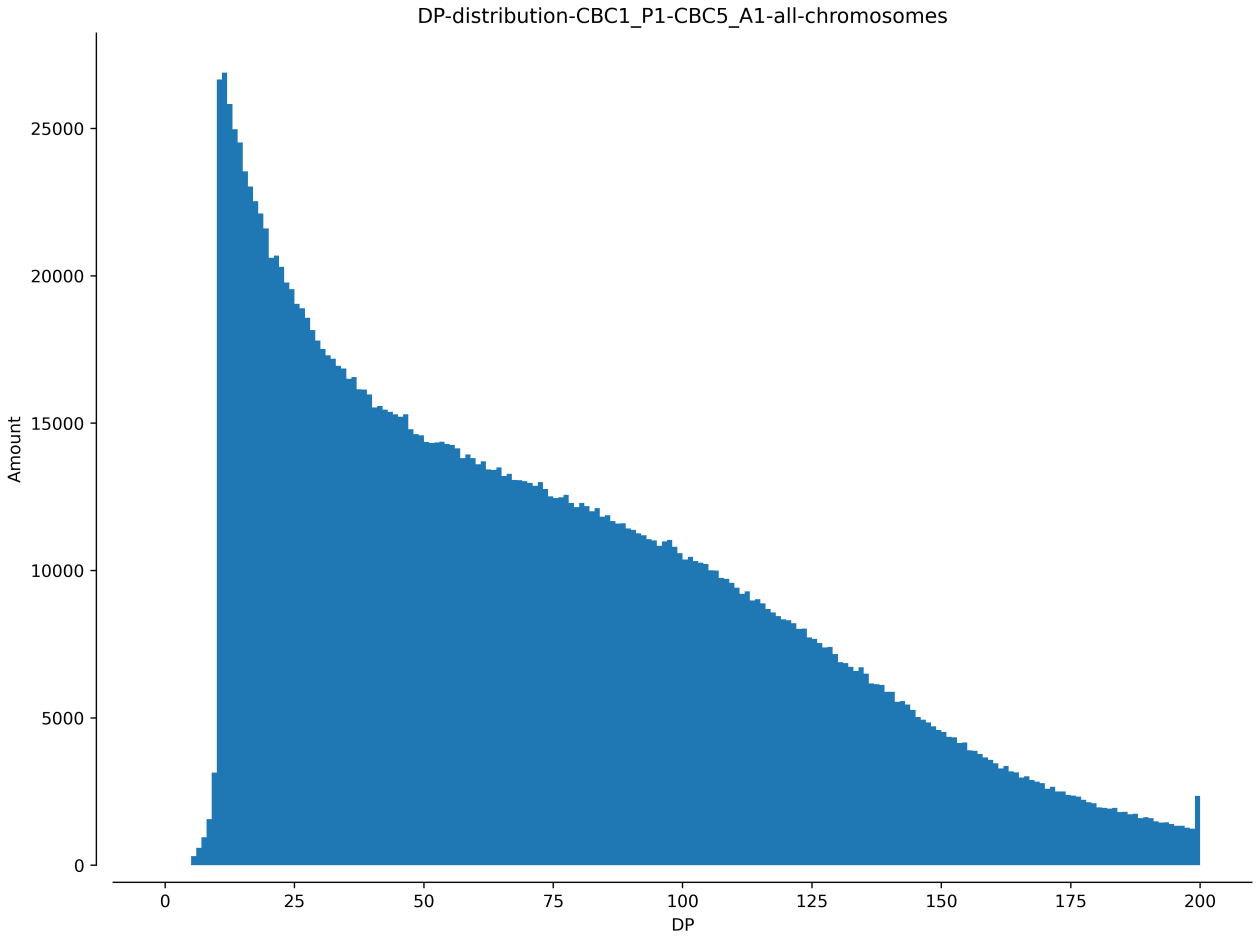
	CHROM	POS	REF	
0	NC_040279.1	912	G	
1	NC_040279.1	948	AGGGGAAAC	A(
2	NC_040279.1	1173	C	
3	NC_040279.1	1390	C	
4	NC_040279.1	1424	T	
...	...	...	...	
1942510	NC_040289.1	41659114	T	

	CHROM	POS	REF
1942511	NC_040289.1	41659137	G
1942512	NC_040289.1	41667130	GTTTCA
1942513	NC_040289.1	41667148	T
1942514	NC_040289.1	41668013	CAGGGTTTAGGGTTTAGGGTTCAGGGTTTAGGGTTTAGGGTTCAGG...

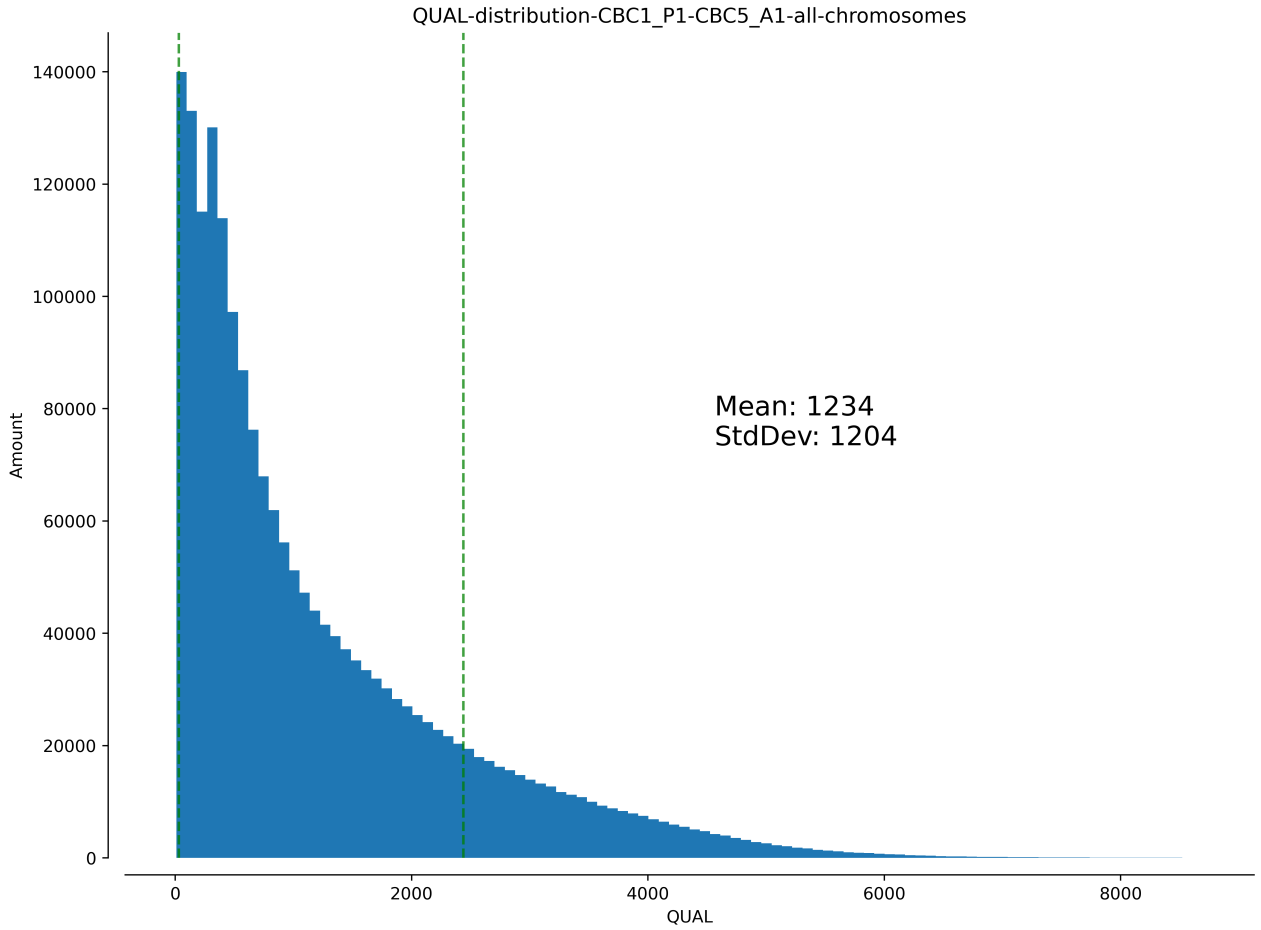
1942515 rows × 14 columns

Verify DP Histogram Cutoff Off by Mean±StdDev

In [32]:
```
plot_variant_hist(samples, vcf_df_01, 'all', 'DP', bins=200, xmax=200)
```



In [33]:
```
plot_variant_hist(samples, vcf_df_01, 'all', 'QUAL', bins=100, MSTD=True)
```



Contingency Table After DP Cutoff by Mean±StdDev

In [34]:
```
contingency_table_2 = contingency_table(samples, vcf_df_01, 'all')
```

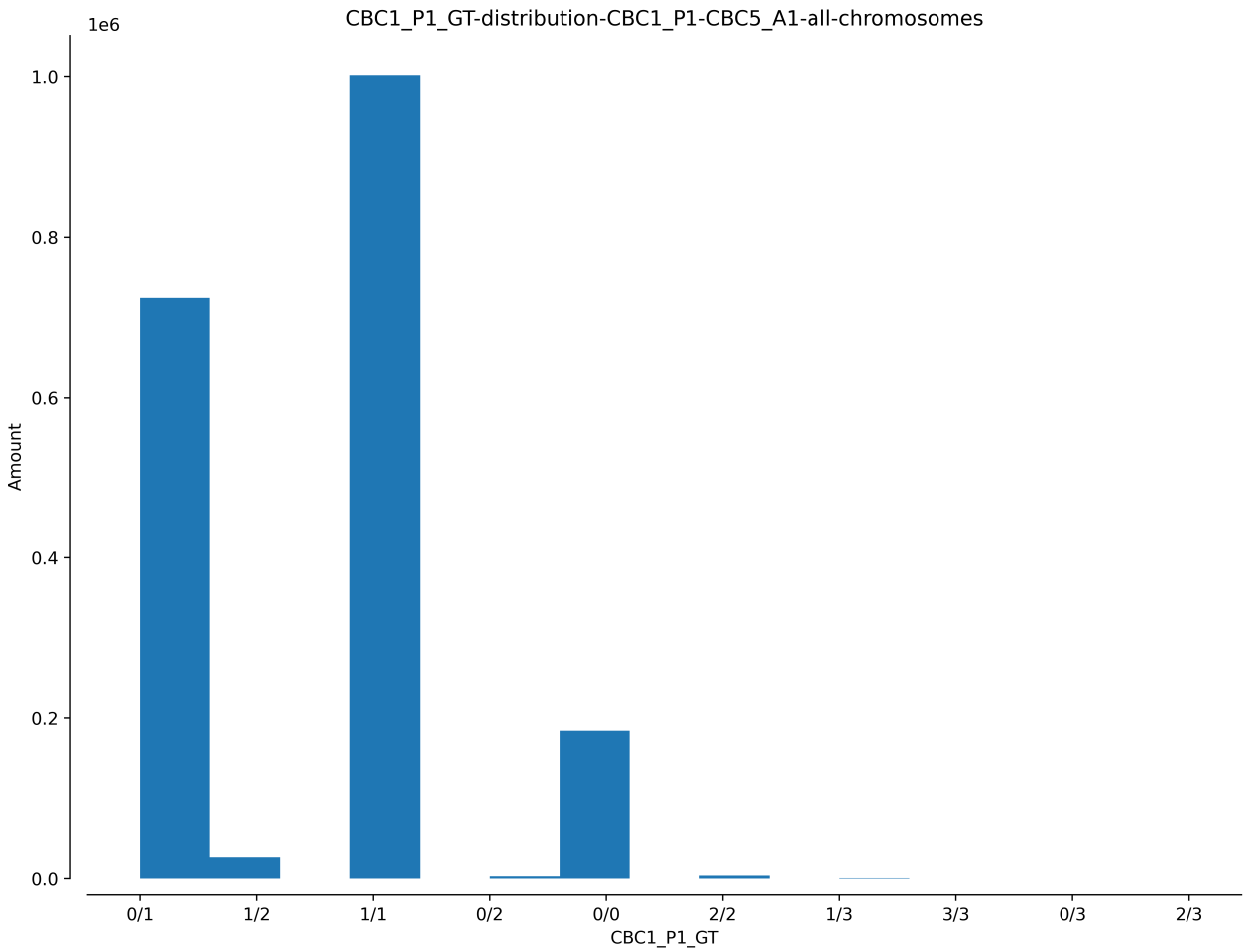
```
Contingency Table - Chromosome all

                CBC5_A1_GT
                       0/0     0/1     1/1  other
CBC1_P1_GT 0/0           0   51039  132478  43507
           0/1      284832  255445  178544  43507
           1/1      211386   19531  765753  43507
           other     43507   43507   43507  43507
```

GT Plot After DP Cutoff by Mean±StdDev

In [35]:
```
# plt.close('all')
# GTplot(samples, vcf_df_01, chrom_len_00)
```

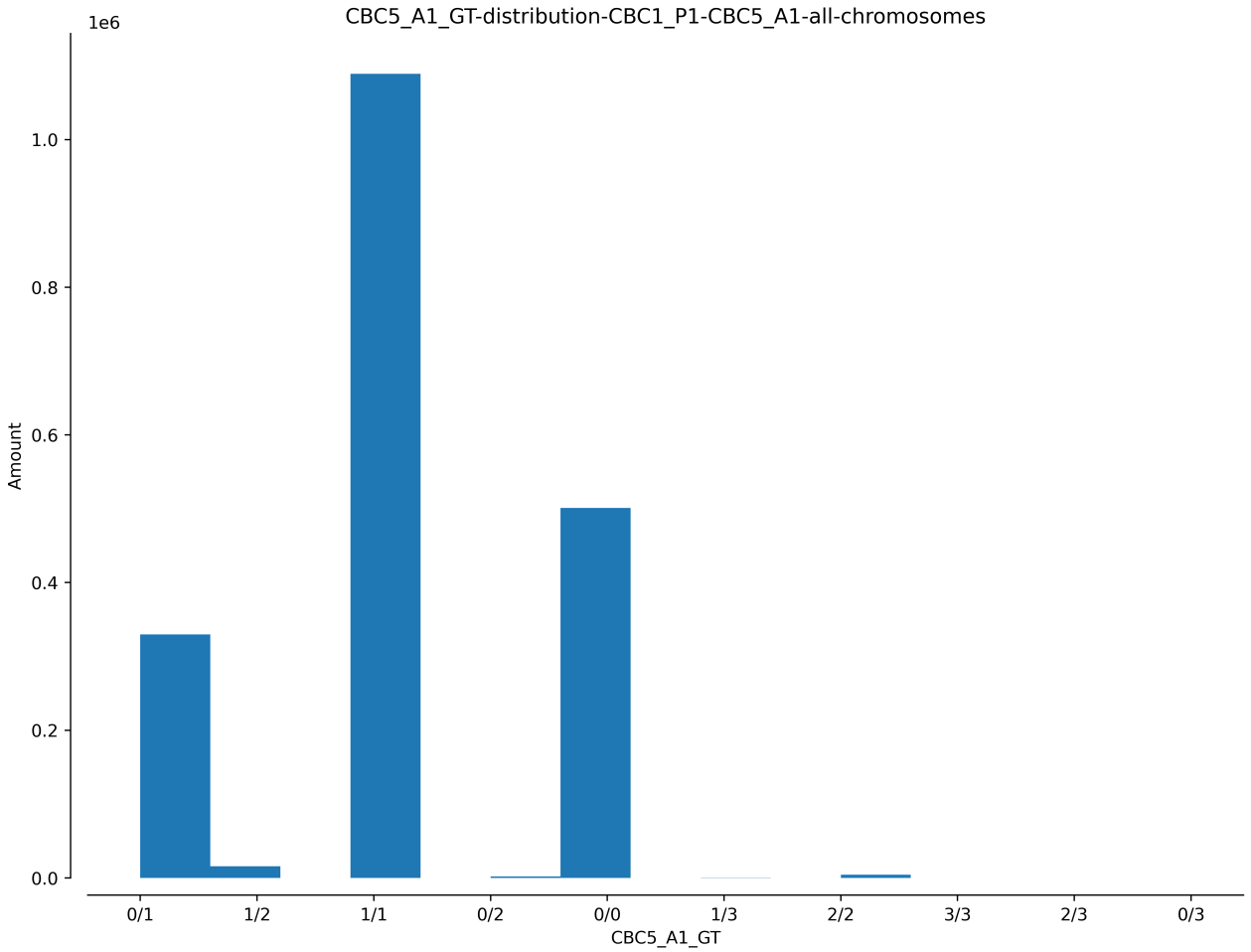
Histogram 'GT' Attribute after DP Cutoff

In [36]:
```
plot_variant_hist(samples, vcf_df_01, 'all', '%s_GT' % progenitor, bins=15)
```



In [37]:

```
plot_variant_hist(samples, vcf_df_01, 'all', '%s_GT' % mutant, bins=15)
```



PART 3: Extract $comp \leq x$ from $TYPE$ Attribute

Extract `complex` TYPE Attribute

In [38]:
```
extract_type = "TYPE != snp, TYPE != mnp, TYPE != ins, TYPE != del"
vcf_df_02 = filter_vcf(vcf_df_01, extract_type)
vcf_df_02
```

Out[38]:

	CHROM	POS	REF	
0	NC_040279.1	1720	GTTTTTTTTTTTTTTTTTTTTTTTATTTCATGGTTCTGTC	GTTTT
1	NC_040279.1	6532	TAAAAAAAAATTAATAAAAAAATTTTAAAAATCGAGGAACTGACAC...	
2	NC_040279.1	8298	AAATGGGCCAAACGAGCCCGAAGACTCGA	
3	NC_040279.1	8523	CAAGCAGACCCGAAA	
4	NC_040279.1	8871	GGCCT	
...	...	...	...	
250809	NC_040289.1	41488977	GTCA	
250810	NC_040289.1	41489075	AG	
250811	NC_040289.1	41489422	AAAAAAAAGGTAAAACAAACTGTGGCTTCCCTGACATGT	
250812	NC_040289.1	41489837	TTAATTTTTTTTTTCTCAG	

	CHROM	POS	REF
250813	NC_040289.1	41658194	CCGC

250814 rows × 14 columns

Examples of `complex` mutation type

In [39]:
```
vcf_df_02[ ['REF', 'ALT'] ].head()
```

Out[39]:

	REF	
0	GTTTTTTTTTTTTTTTTTTTTTTTATTTCATGGTTCTGTC	GTTTTTTTTTTTATTTTTTTTTTATTTCATC
1	TAAAAAAAAATTAATAAAAAAATTTTAAAAATCGAGGAACTGACAC...	TAAAAAAAAAATAA
2	AAATGGGCCAAACGAGCCCGAAGACTCGA	AAACGGGCCAAACGAGCCCGA
3	CAAGCAGACCCGAAA	CAGGCAC
4	GGCCT	

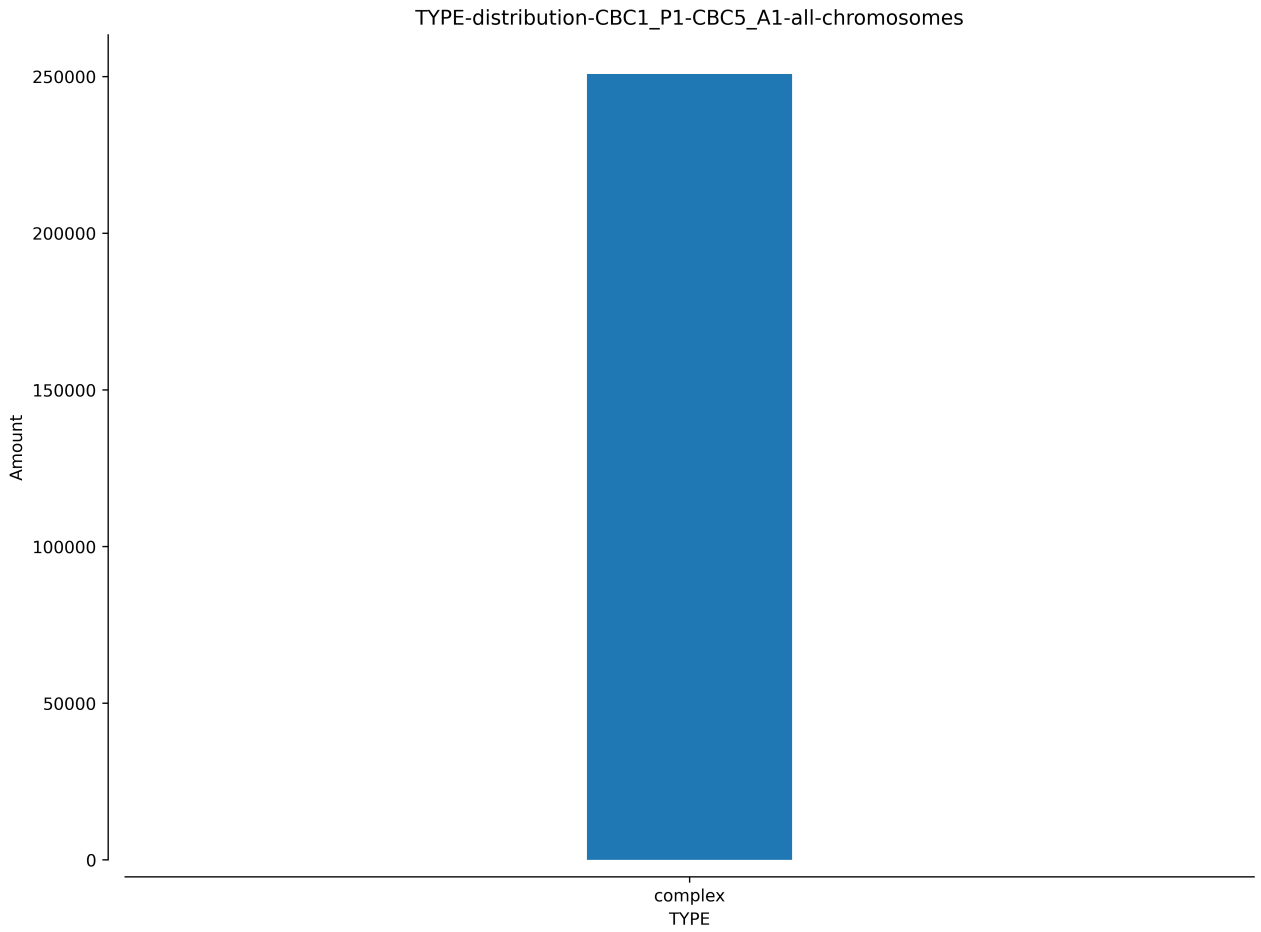
In [40]:
```
vcf_df_02[ ['REF', 'ALT'] ].tail()
```

Out[40]:

	REF	ALT
250809	GTCA	TTCG
250810	AG	CA
250811	AAAAAAAAGGTAAAACAAACTGTGGCTTCCCTGACATGT	AAAAAAAAA
250812	TTAATTTTTTTTTTCTCAG	ATATTTTTTTTTTCTCAG
250813	CCGC	TCAT

TYPE `complex` Histogram Verification

In [41]:
```
plot_variant_hist(samples, vcf_df_02, 'all', 'TYPE', bins=5)
```



Contingency Table - `complex` TYPE only

In [42]:
```
contingency_table_3 = contingency_table(samples, vcf_df_02, 'all')
```

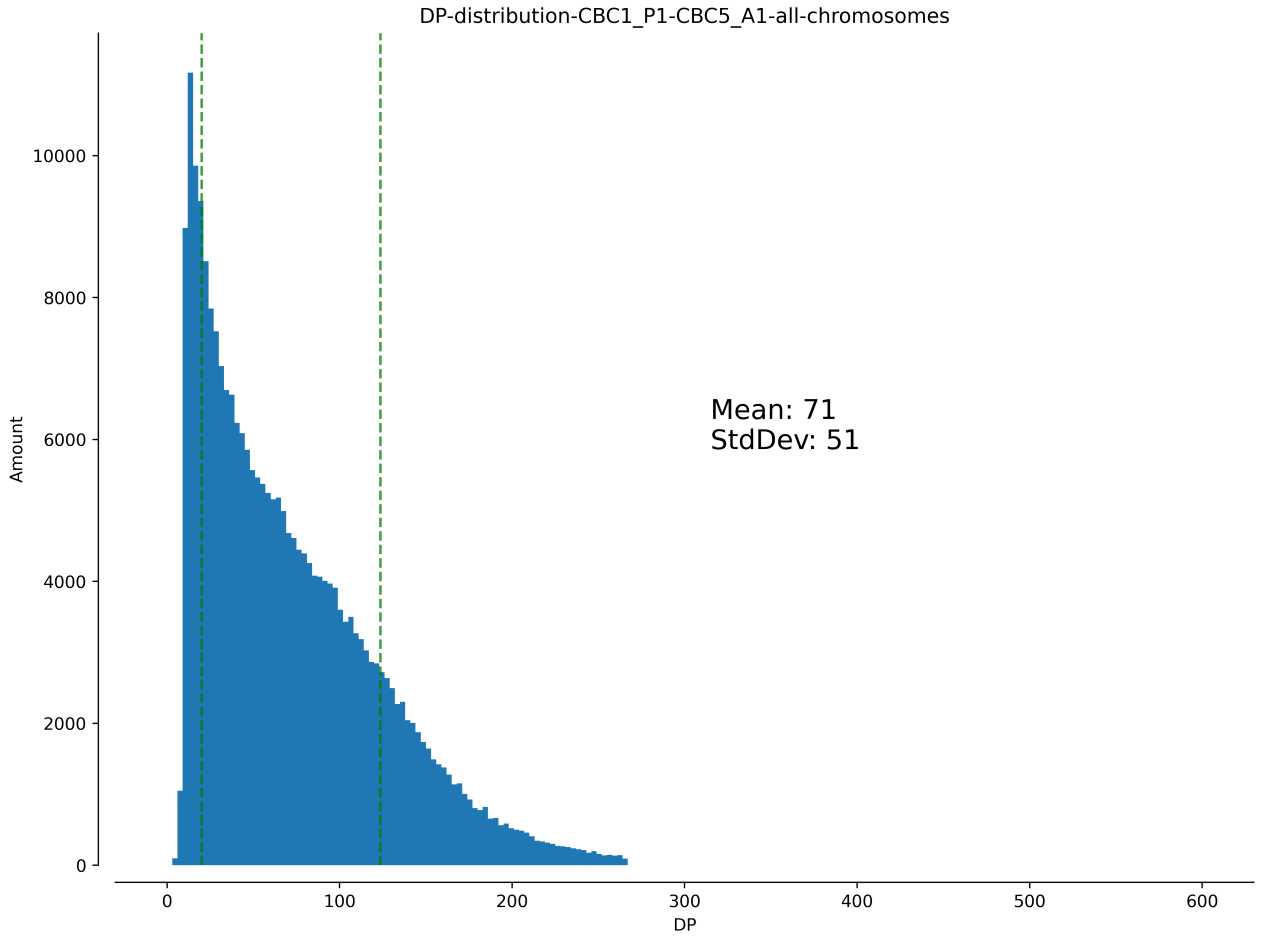
```
Contingency Table - Chromosome all

                CBC5_A1_GT
                       0/0    0/1    1/1  other
CBC1_P1_GT 0/0           0   8142  13940  26231
           0/1       35574  38434  19985  26231
           1/1       23907   2145  82456  26231
           other     26231  26231  26231  26231
```

GT Plot - `complex` TYPE only

In [43]:
```
# plt.close('all')
# GTplot(samples, vcf_df_02, chrom_len_00)
```

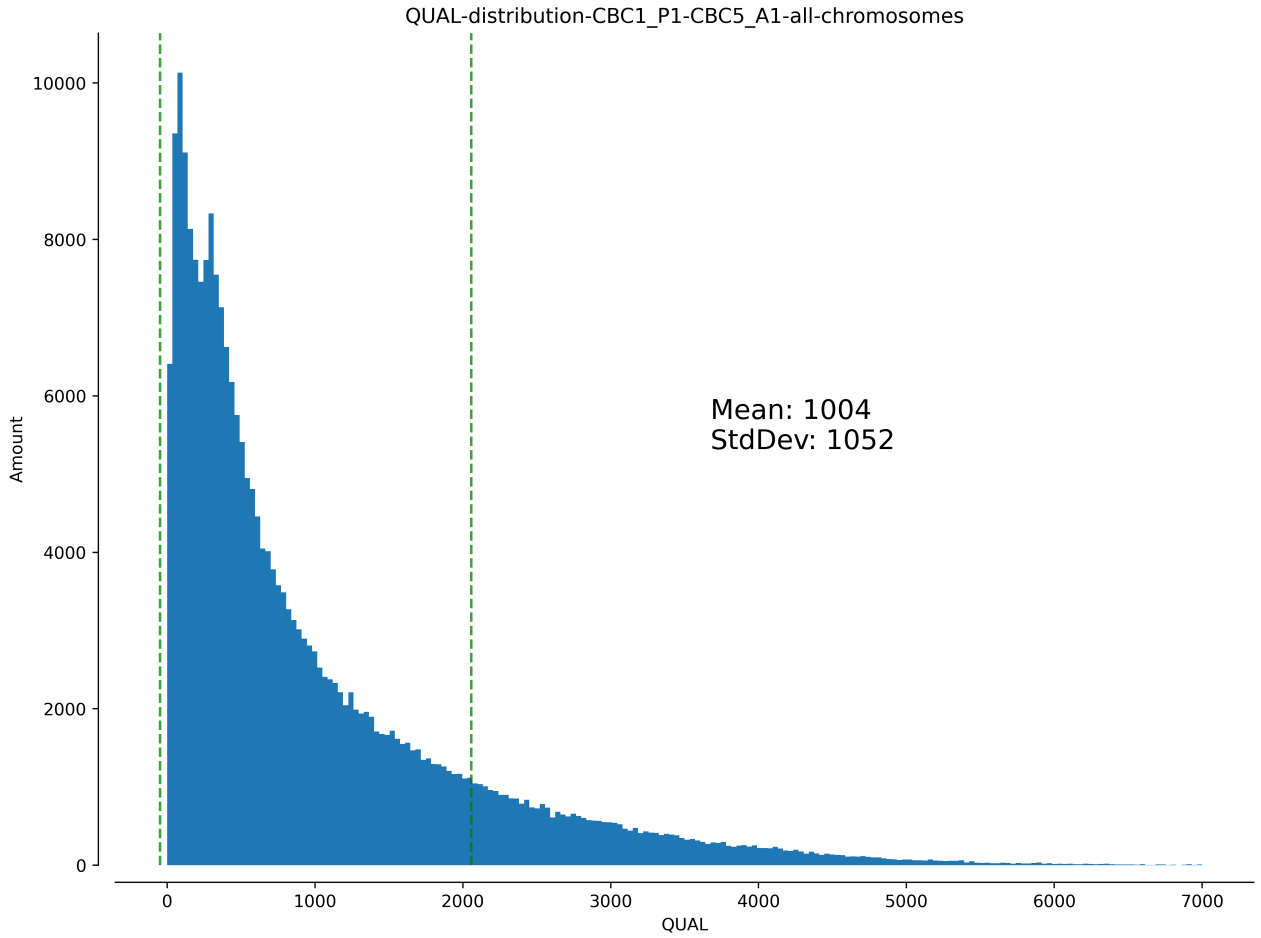
Histograms - DP, QUAL, and GT Attributes after TYPE Filtering

In [44]:
```
plot_variant_hist(samples, vcf_df_02, 'all', 'DP', bins=200, MSTD=True, xmax=600
```



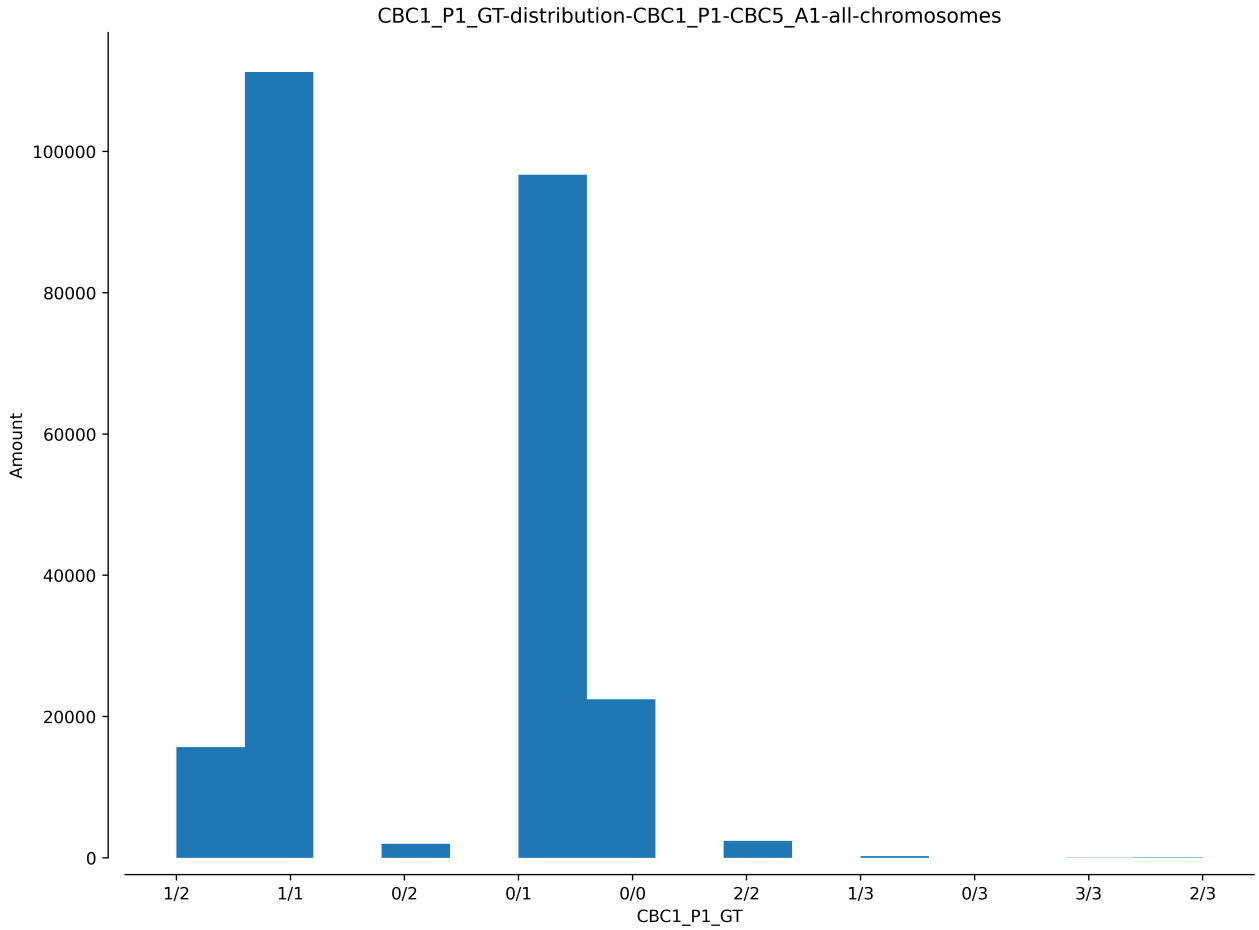
In [45]:

```
plot_variant_hist(samples, vcf_df_02, 'all', 'QUAL', bins=200,  MSTD=True, xmax=
```



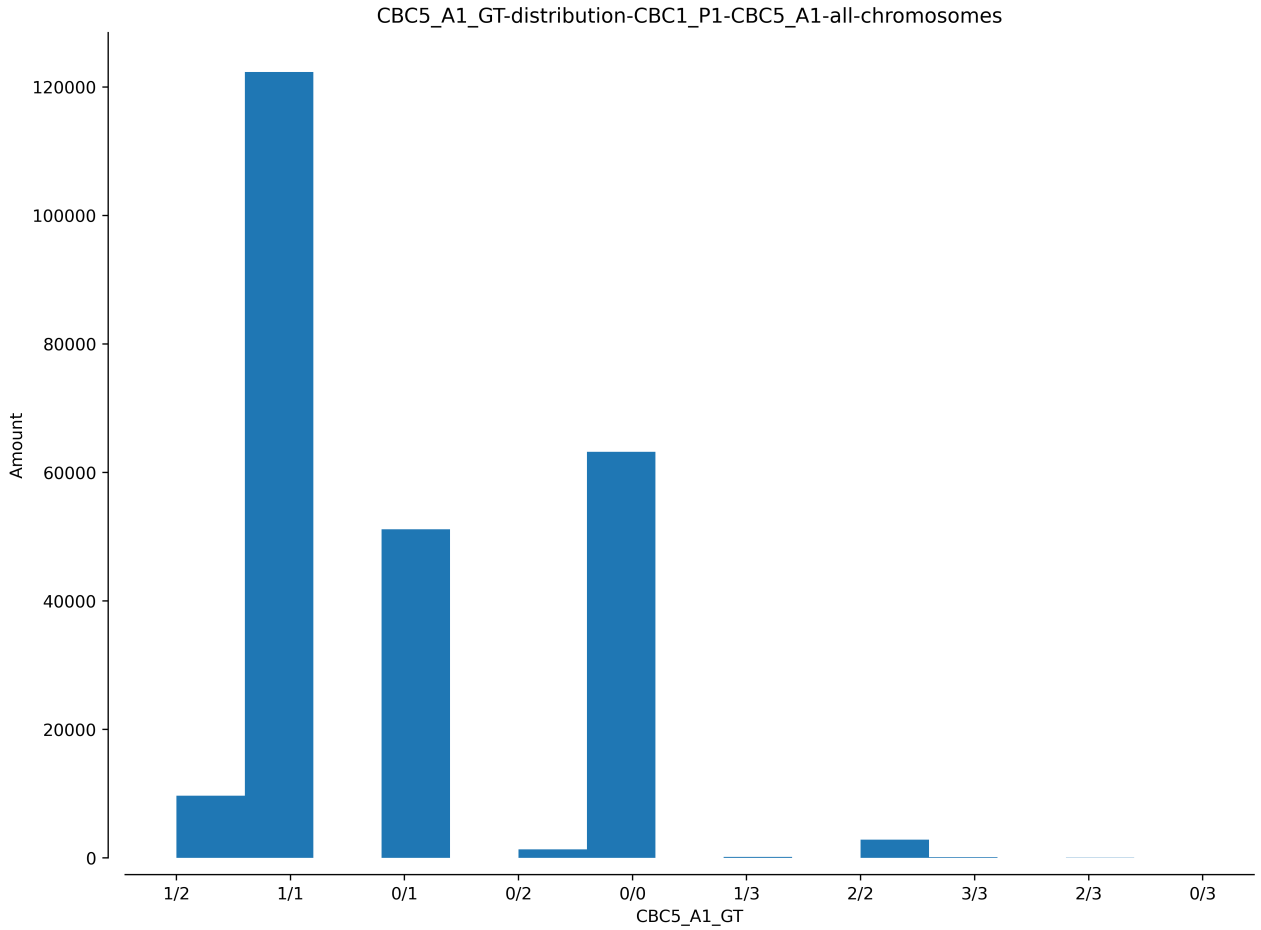
In [46]:

```
plot_variant_hist(samples, vcf_df_02, 'all', '%s_GT' % progenitor, bins=15)
```



In [47]:

```
plot_variant_hist(samples, vcf_df_02, 'all', '%s_GT' % mutant, bins=15)
```



PART 4: Cutting Off by Mean±StdDev Histograms of $QUAL$ Attribute

In [48]:
```
# cutoff_left = vcf_df_02.QUAL.mean() - vcf_df_02.QUAL.std()
# cutoff_right = vcf_df_02.QUAL.mean() + vcf_df_02.QUAL.std()

# filter_qual = "QUAL >= %i, QUAL <= %i" % (cutoff_left, cutoff_right)
# print(filter_qual)

# vcf_df_03 = filter_vcf(vcf_df_02, filter_qual)
# vcf_df_03
```

Verify `DP` and `QUAL` Histograms after `QUAL` Cutoff Off by Mean±StdDev

In [49]:
```
# plot_variant_hist(samples, vcf_df_03, 'all', 'DP', bins=200, xmax=200)
```

In [50]:
```
# plot_variant_hist(samples, vcf_df_03, 'all', 'QUAL', bins=100, xmax=3500)
```

Contingency Table After QUAL Cutoff by Mean±StdDev

In [51]:
```
# contingency_table_4 = contingency_table(samples, vcf_df_03, 'all')
# contingency_table_4
```

GT Plot After QUAL Cutoff by Mean±StdDev

In [52]:
```
# plt.close('all')
# GTplot(samples, vcf_df_03, chrom_len_00)
```

Histograms after QUAL Cutoff by Mean±StdDev

In [53]:
```
# plot_variant_hist(samples, vcf_df_03, 'all', 'CBC1_P1_GT', bins=9)
```

In [54]:
```
# plot_variant_hist(samples, vcf_df_03, 'all', 'CBC5_A1_GT', bins=9)
```

PART 5: Filtering GTs 0/0, 1/1, 'Other'

Filter out where samples GTs are the same (0/0, 1/1) and have 'Other'

In [55]:
```
progenitor_gts_filter = "CBC1_P1_GT != ./., CBC1_P1_GT != 0/2, CBC1_P1_GT != 1/2
vcf_df_04 = filter_vcf(vcf_df_02, progenitor_gts_filter)

mutant_gts_filter = "CBC5_A1_GT != ./., CBC5_A1_GT != 0/2, CBC5_A1_GT != 1/2, CB
vcf_df_04 = filter_vcf(vcf_df_04, mutant_gts_filter)

genotypes = ['0/0', '1/1']
for genotype in genotypes:
    vcf_df_04 = filter_similar_gt(samples, vcf_df_04, genotype)

vcf_df_04
```

Out[55]:

	CHROM	POS	REF	ALT	
0	NC_040279.1	8871	GGCCT	AGCCC	1
1	NC_040279.1	12210	ATC	CAA	
2	NC_040279.1	12243	GAC	AAA	1
3	NC_040279.1	14771	GGTT	AGTC	
4	NC_040279.1	60920	GA	AG	1
...	...	...	...	...	
142122	NC_040289.1	41487815	AAAC	CAAT	2
142123	NC_040289.1	41488049	TTTT	ATTG	3
142124	NC_040289.1	41488526	TTATTTT	TTGTC	1
142125	NC_040289.1	41489422	AAAAAAAAGGTAAAACAAACTGTGGCTTCCCTGACATGT	AAAAAAAAA	
142126	NC_040289.1	41658194	CCGC	TCAT	

142127 rows × 14 columns

Contingency Table after GT Filtering

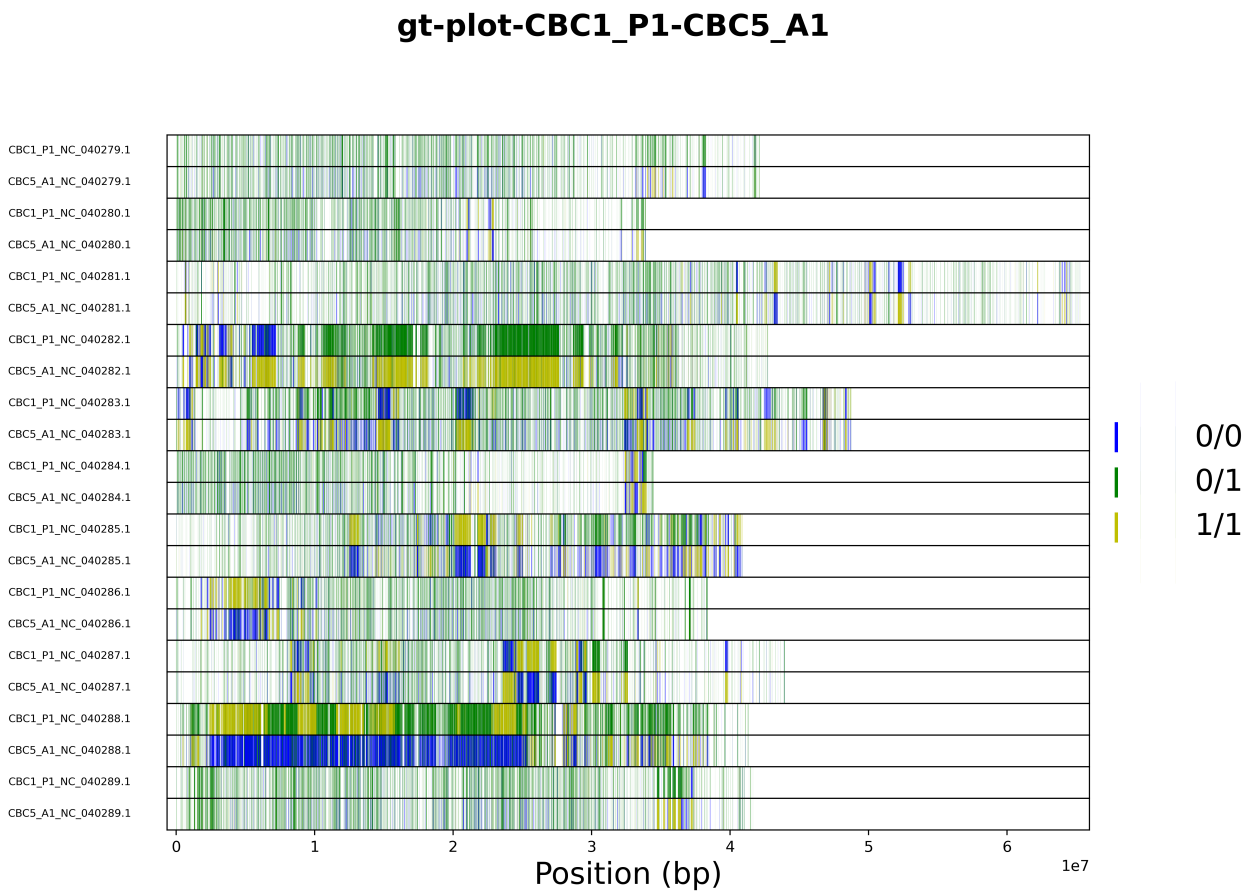
In [56]:
```
contingency_table_5 = contingency_table(samples, vcf_df_04, 'all')
```

```
Contingency Table - Chromosome all

                  CBC5_A1_GT
                         0/0    0/1    1/1 other
CBC1_P1_GT 0/0             0   8142  13940     0
           0/1         35574  38434  19985     0
           1/1         23907   2145      0     0
           other           0      0      0     0
```

GT Plot after GT Filtering

In [57]:

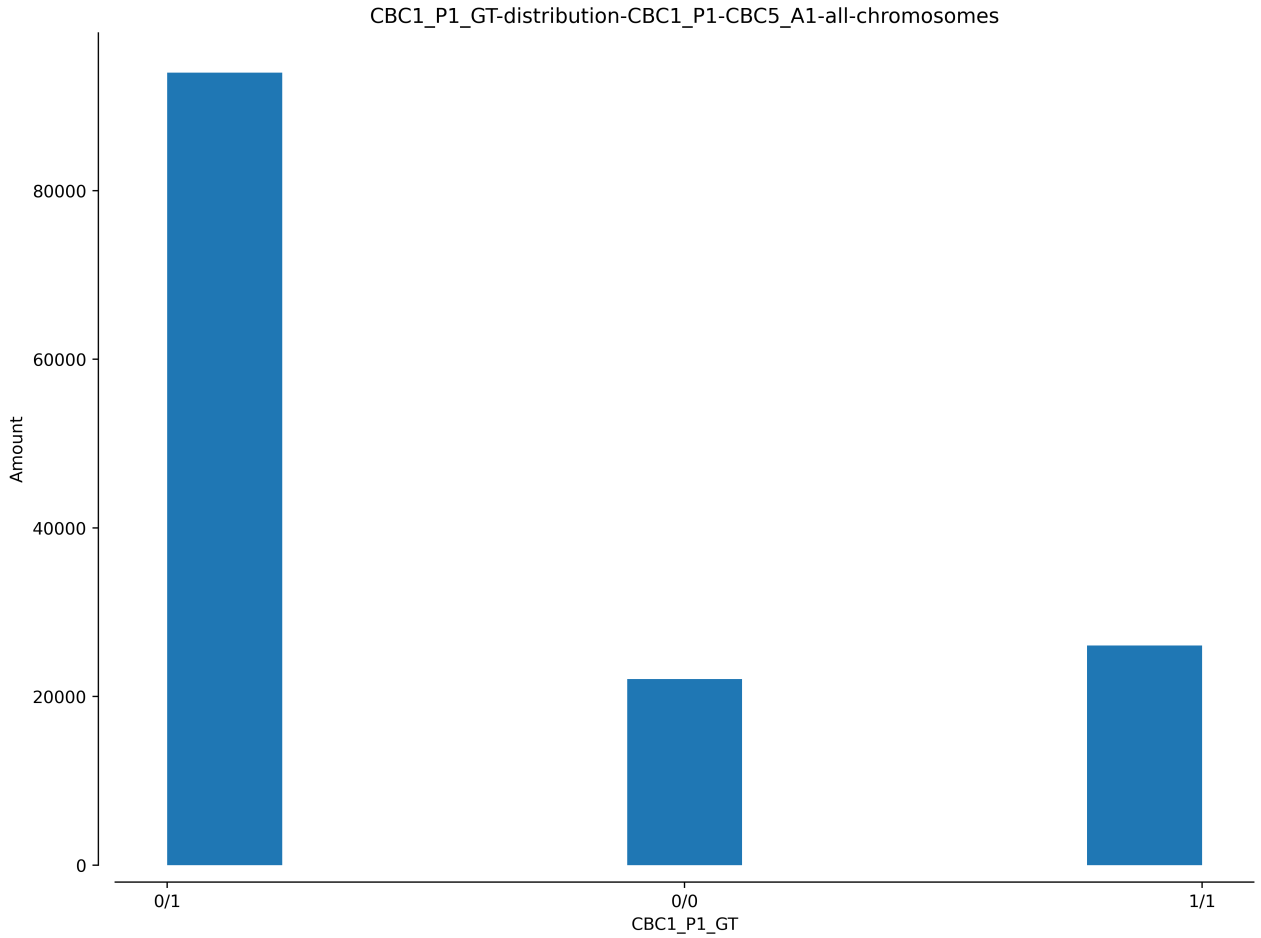
```
plt.close('all')
GTplot(samples, vcf_df_04, chrom_len_00)
```



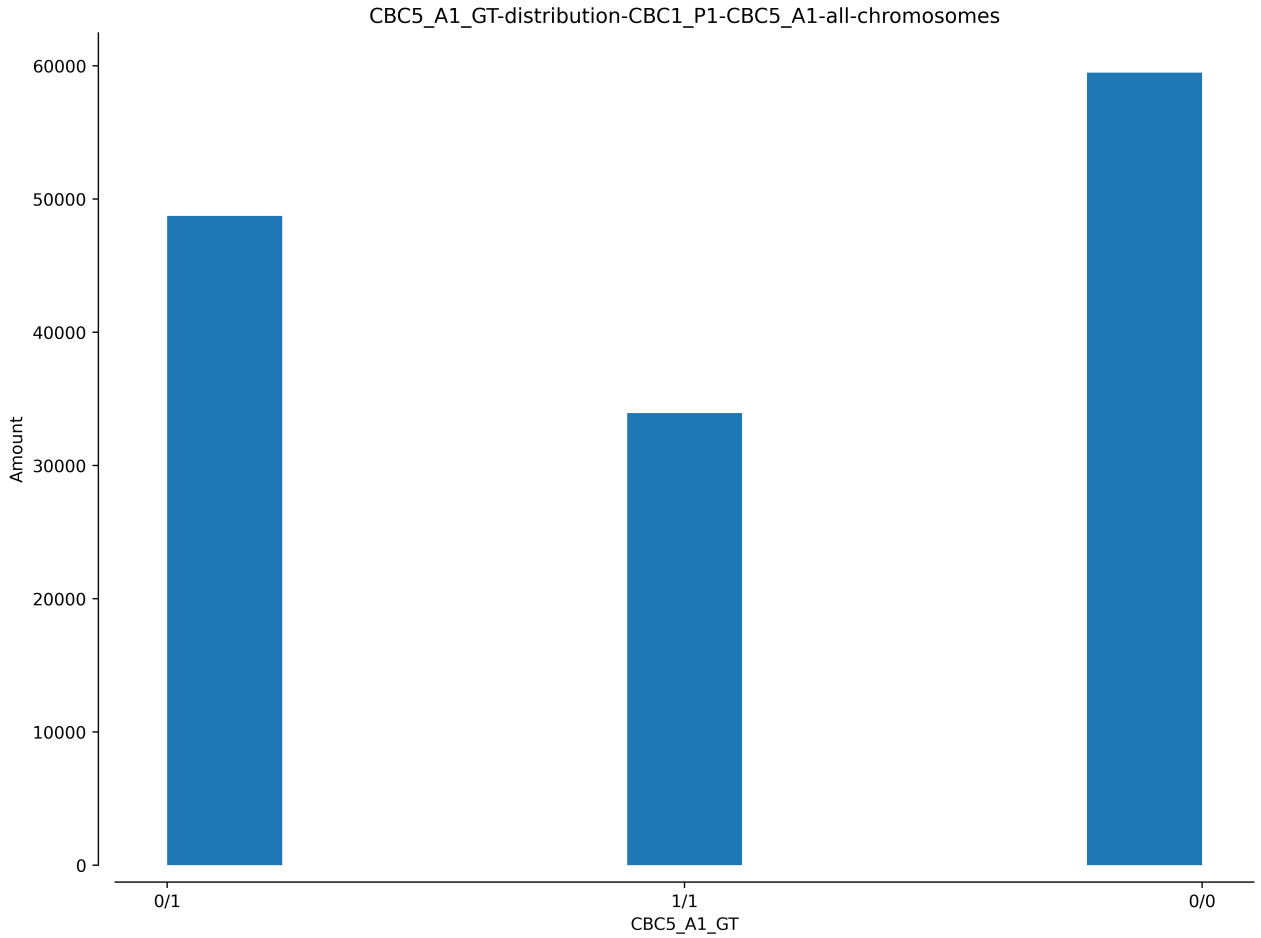
Histograms GT after GT Filtering

In [58]:

```
plot_variant_hist(samples, vcf_df_04, 'all', '%s_GT' % progenitor, bins=9)
```



In [59]:
```
plot_variant_hist(samples, vcf_df_04, 'all', '%s_GT' % mutant, bins=9)
```



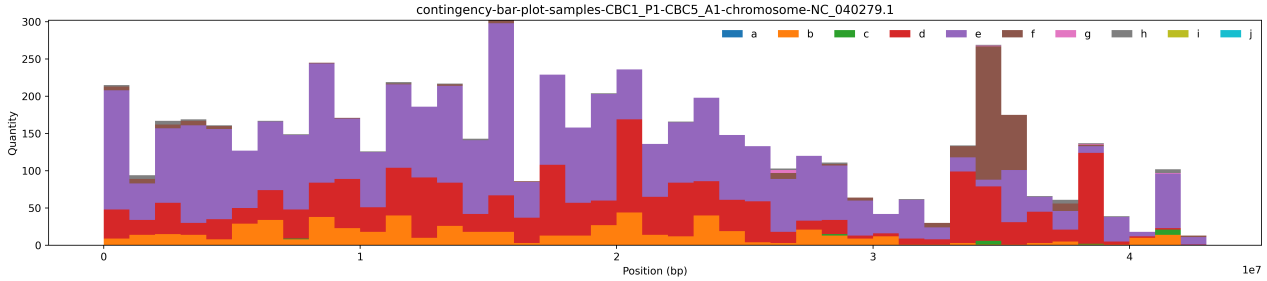
PART 6: Stacked Bar Plots

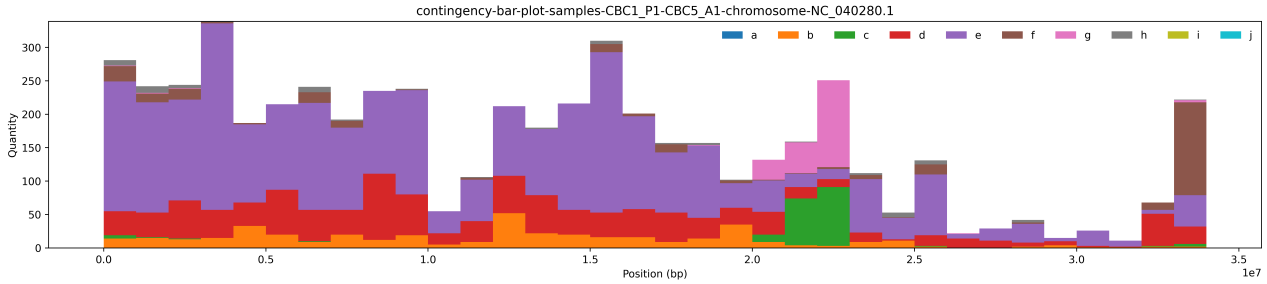
In [60]:
```
ct_guide()
```

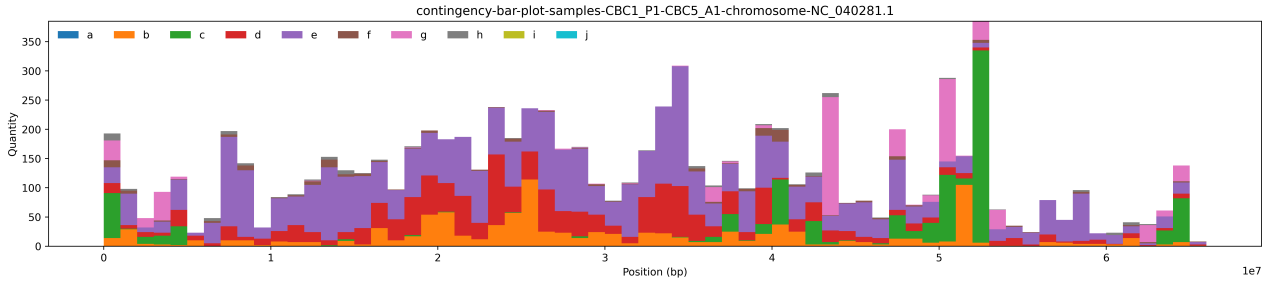
Out[60]:

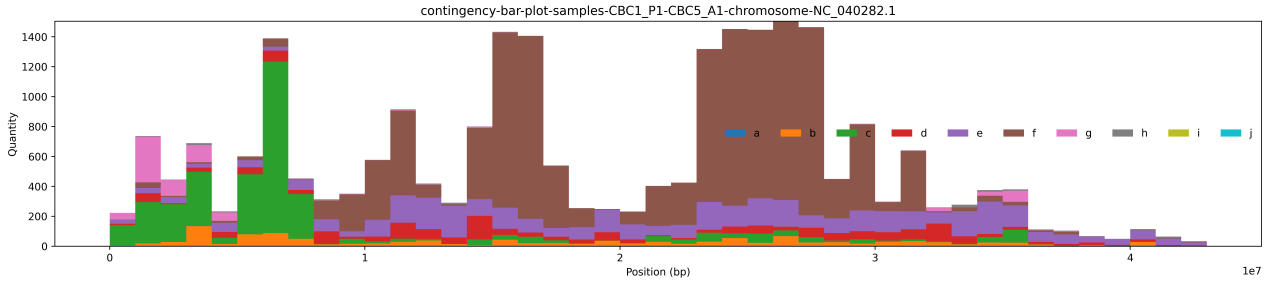
		Mutant			
		0/0	**0/1**	**1/1**	**other**
Progenitor	**0/0**	a	b	c	
	0/1	d	e	f	
	1/1	g	h	i	
	other				j

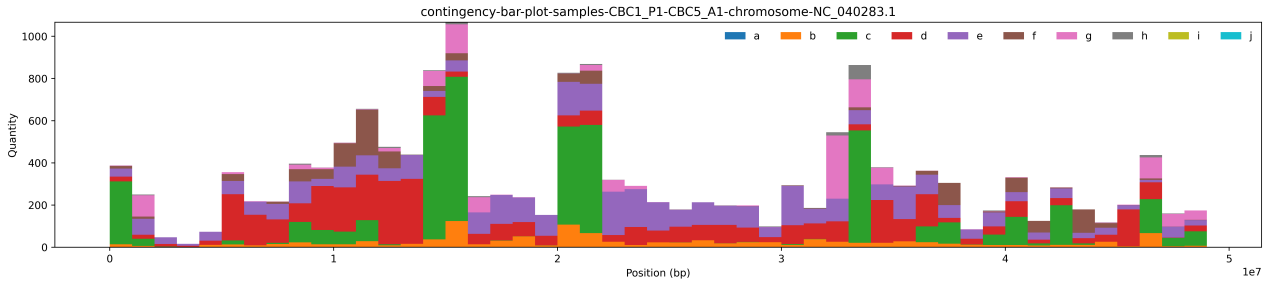
In [61]:
```
plt.close('all')
window_size = 1000000
CTbarPlots(samples, vcf_df_04, chrom_len_00, window_size)
```

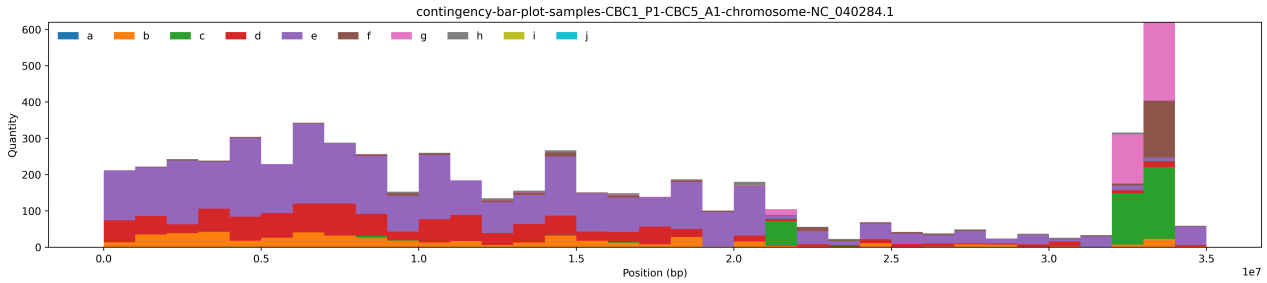

contingency-bar-plot-samples-CBC1_P1-CBC5_A1-chromosome-NC_040279.1
a b c d e f g h i j
Quantity
Position (bp)
1e7

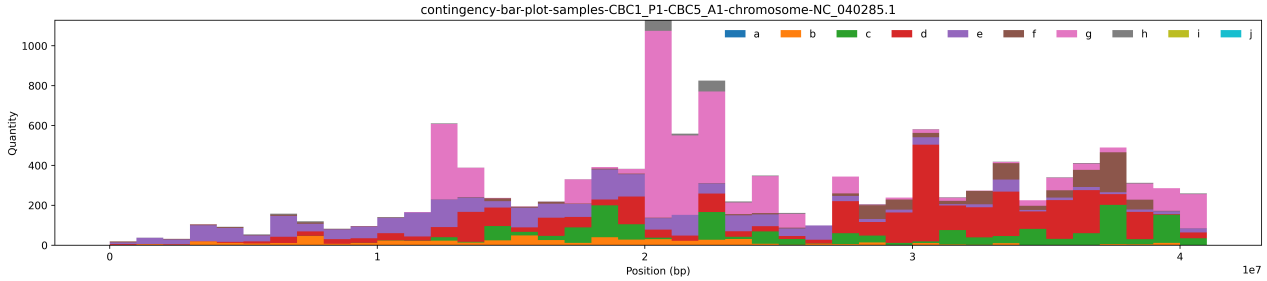

contingency-bar-plot-samples-CBC1_P1-CBC5_A1-chromosome-NC_040280.1
a b c d e f g h i j
Quantity
Position (bp)
1e7

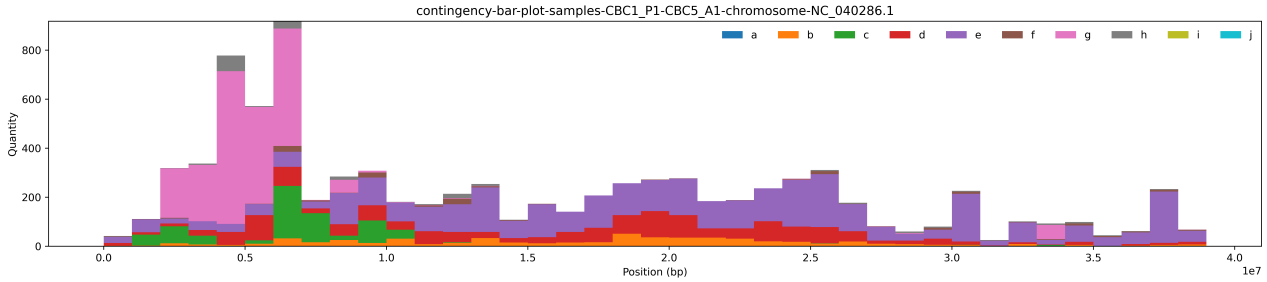

contingency-bar-plot-samples-CBC1_P1-CBC5_A1-chromosome-NC_040281.1
a b c d e f g h i j
Quantity
Position (bp)
1e7

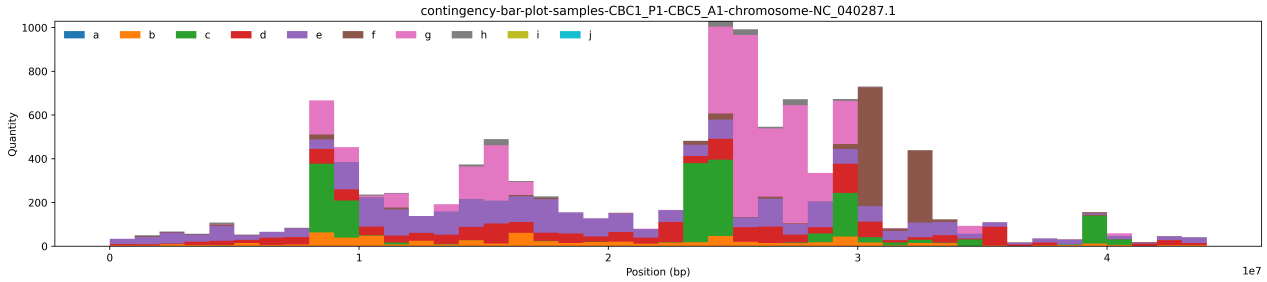

contingency-bar-plot-samples-CBC1_P1-CBC5_A1-chromosome-NC_040282.1
a b c d e f g h i j
Quantity
Position (bp)
1e7

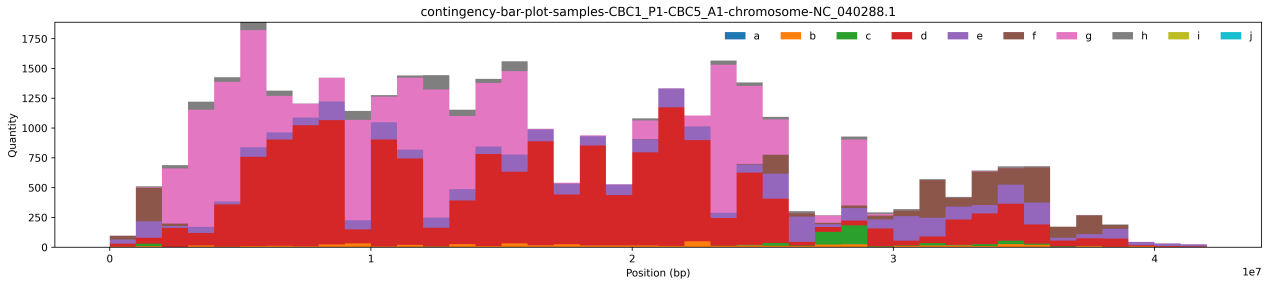

contingency-bar-plot-samples-CBC1_P1-CBC5_A1-chromosome-NC_040283.1
a b c d e f g h i j
Quantity
Position (bp)
1e7

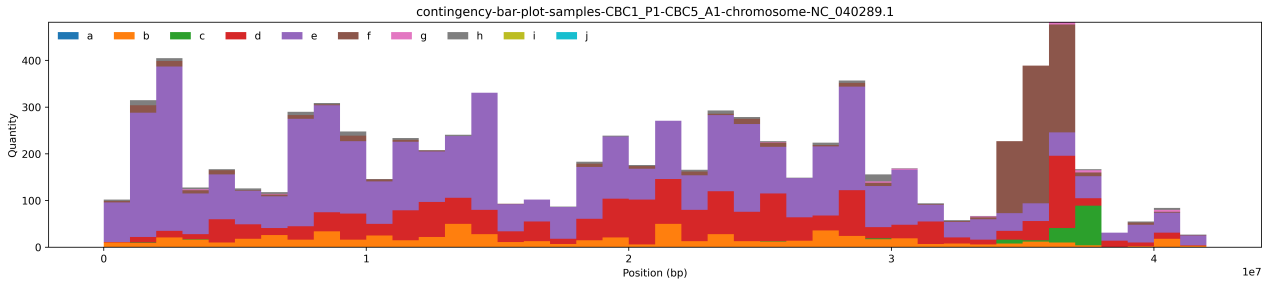

contingency-bar-plot-samples-CBC1_P1-CBC5_A1-chromosome-NC_040284.1
a b c d e f g h i j
Quantity
Position (bp)
1e7







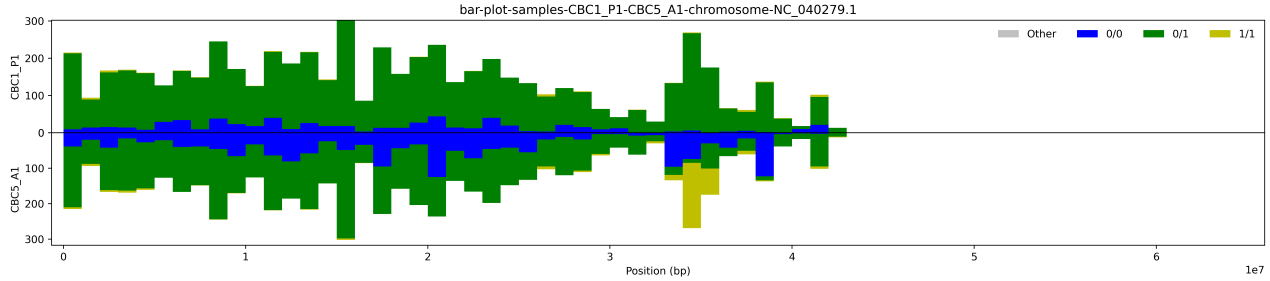


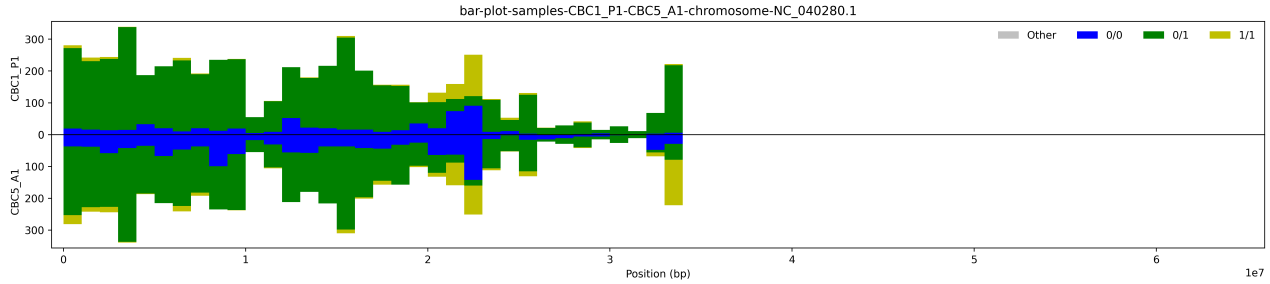


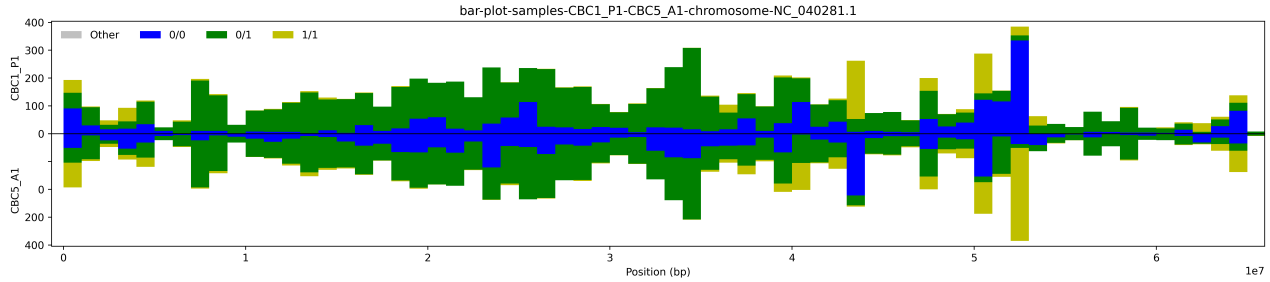
PART 7: Bar Plots per Chromosome

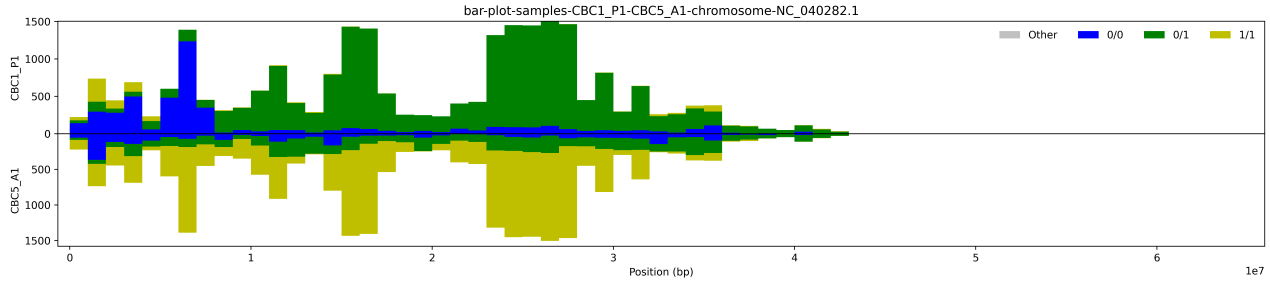
In [62]:

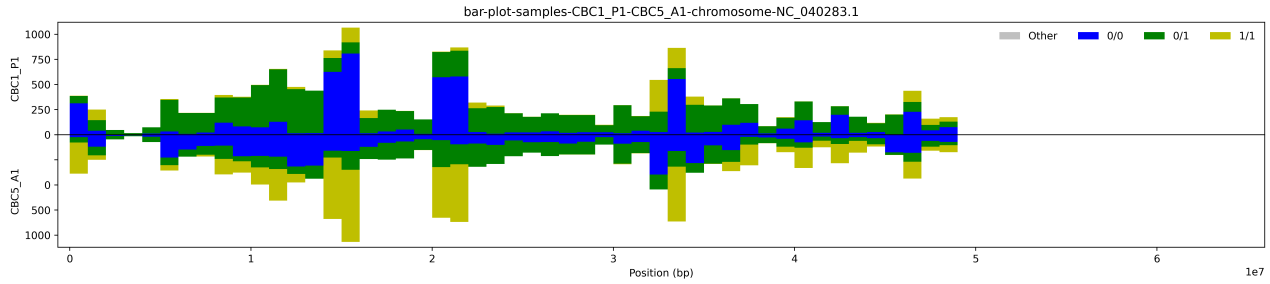
```
# suppress all the warnings from the inverted tickes of bar plots
import warnings
warnings.filterwarnings('ignore')

plt.close('all')
GTbarPlots(samples, vcf_df_04, chrom_len_00, window_size)
```

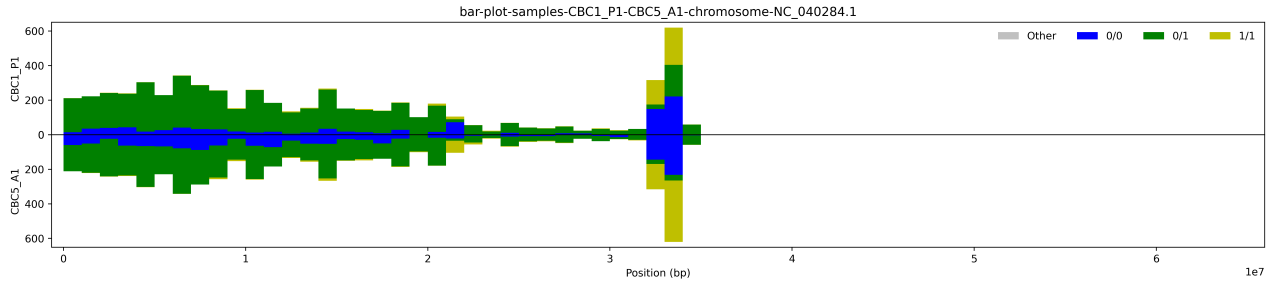

bar-plot-samples-CBC1_P1-CBC5_A1-chromosome-NC_040279.1
Other
0/0
0/1
1/1
CBC1_P1
CBC5_A1
Position (bp)
1e7

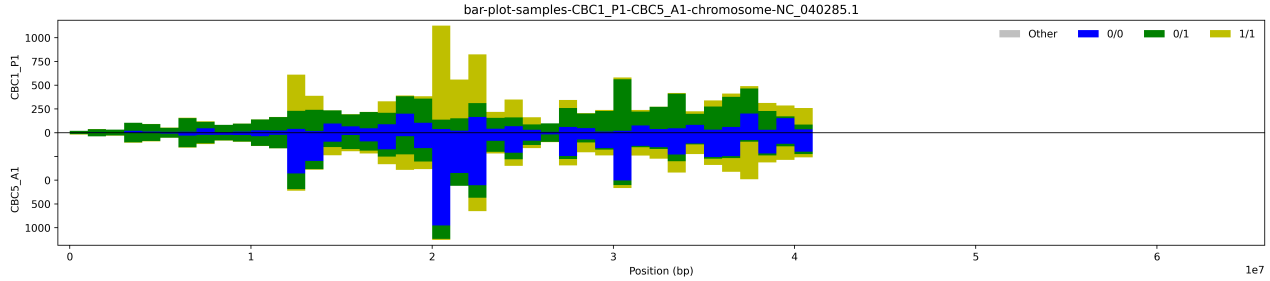

bar-plot-samples-CBC1_P1-CBC5_A1-chromosome-NC_040280.1
Other
0/0
0/1
1/1
CBC1_P1
CBC5_A1
Position (bp)
1e7

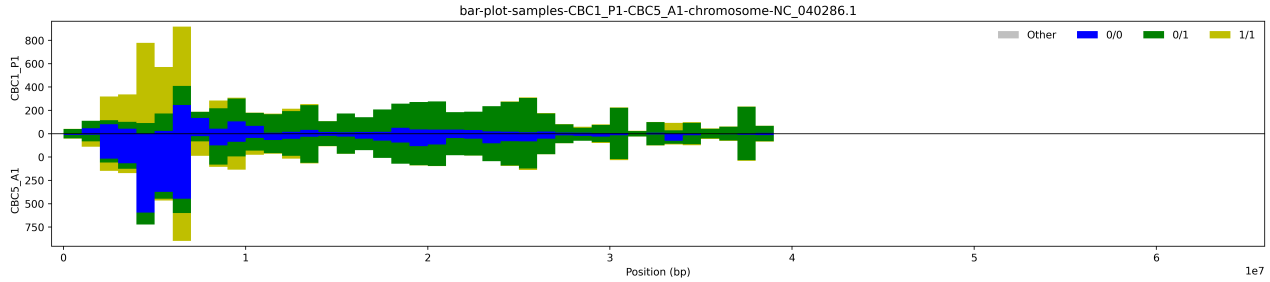

bar-plot-samples-CBC1_P1-CBC5_A1-chromosome-NC_040281.1
Other
0/0
0/1
1/1
CBC1_P1
CBC5_A1
Position (bp)
1e7

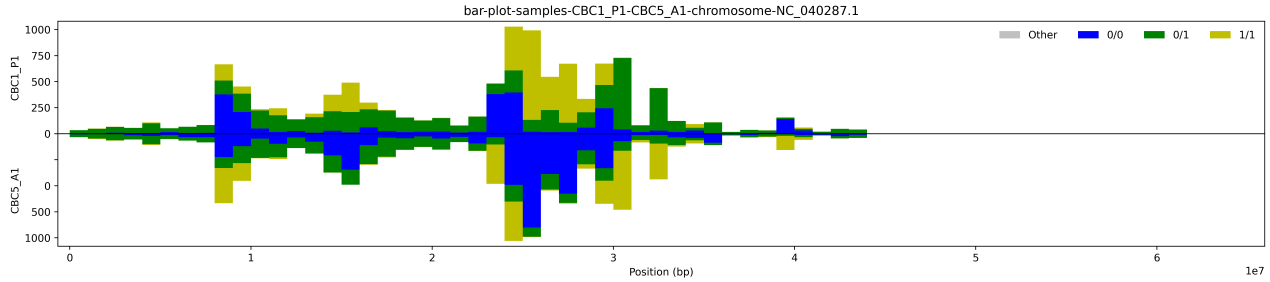

bar-plot-samples-CBC1_P1-CBC5_A1-chromosome-NC_040282.1
Other
0/0
0/1
1/1
CBC1_P1
CBC5_A1
Position (bp)
1e7

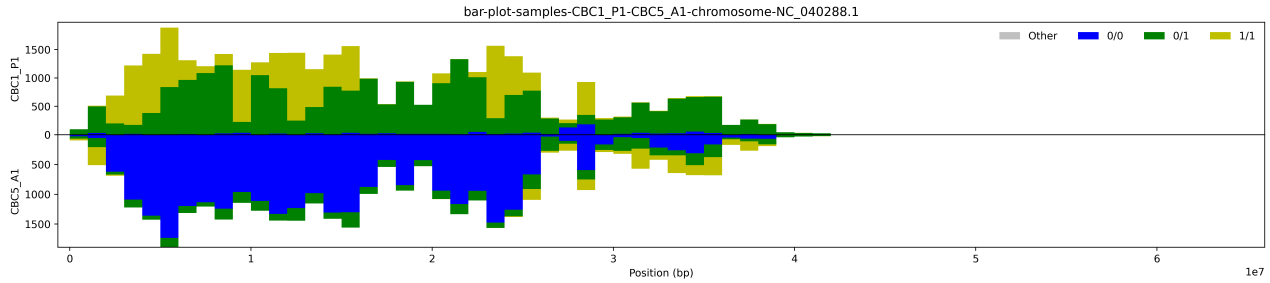

bar-plot-samples-CBC1_P1-CBC5_A1-chromosome-NC_040283.1
Other
0/0
0/1
1/1
CBC1_P1
CBC5_A1
Position (bp)
1e7

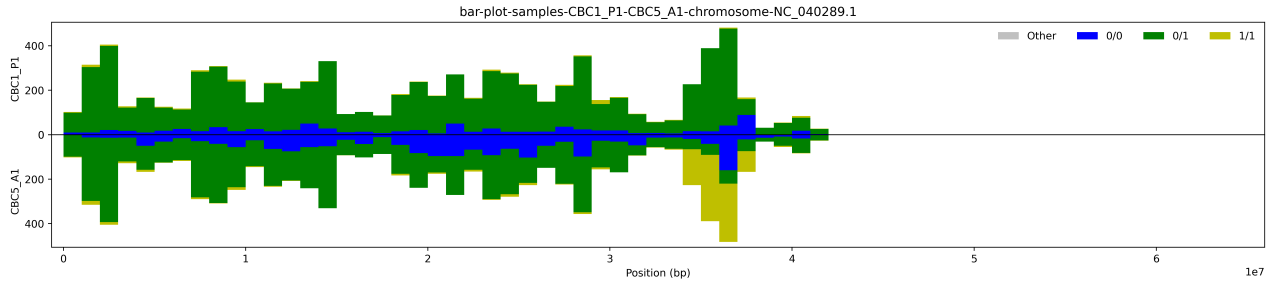

bar-plot-samples-CBC1_P1-CBC5_A1-chromosome-NC_040284.1
Other
0/0
0/1
1/1
CBC1_P1
CBC5_A1
Position (bp)
1e7







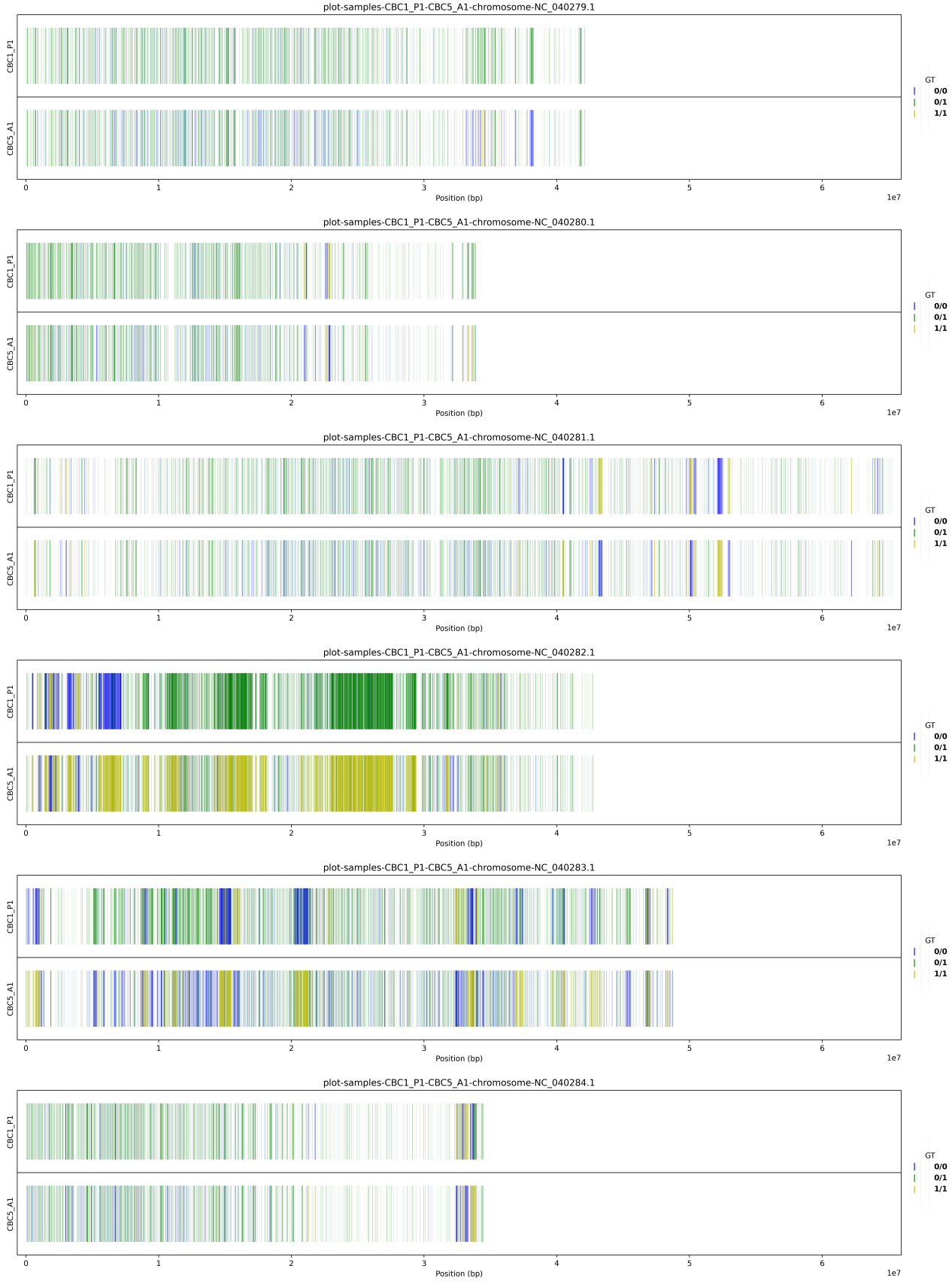


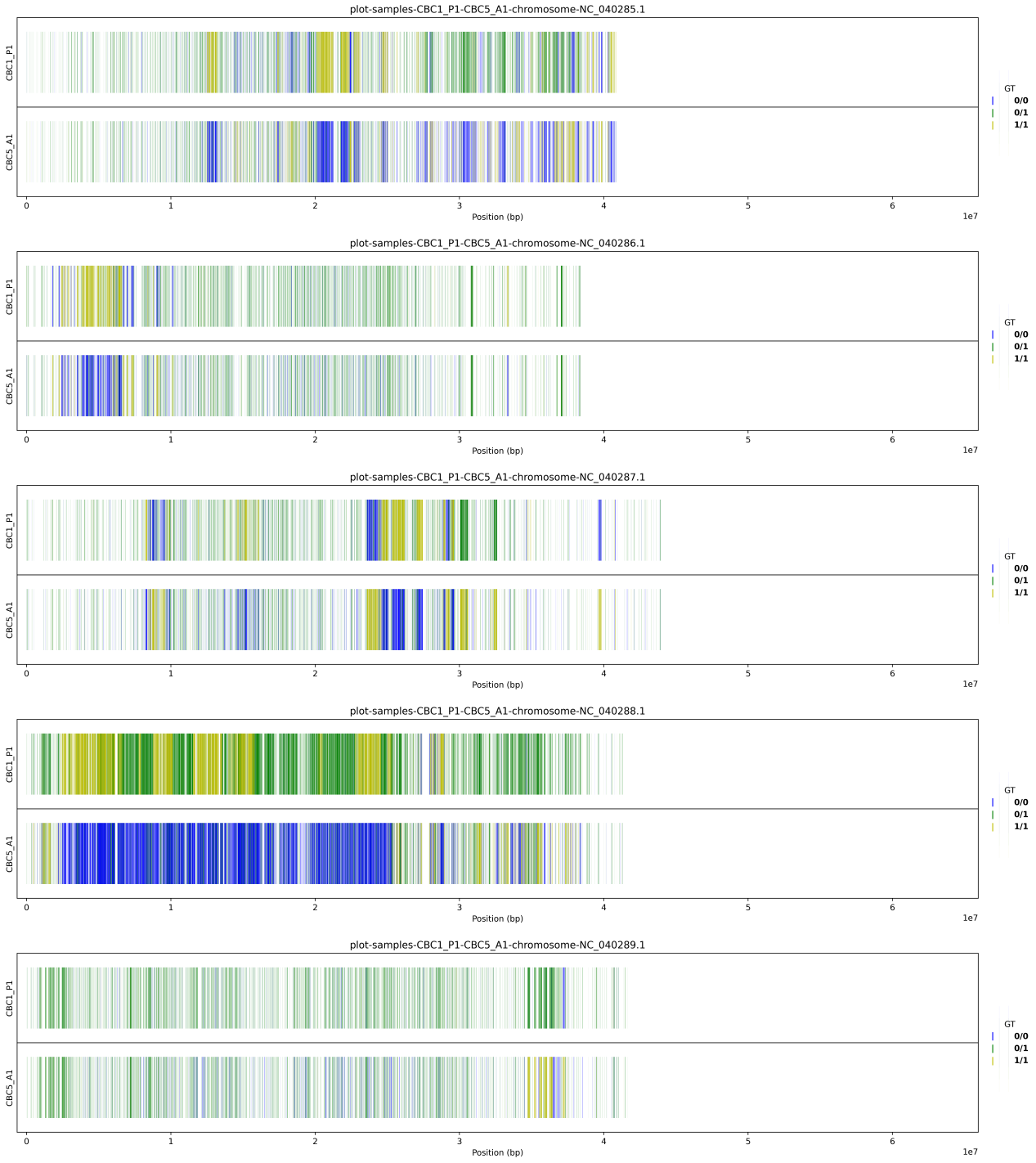


PART 8: GT Plots per Chromosome

In [63]:

```
plt.close('all')
GTplots(samples, vcf_df_04, chrom_len_00)
```


plot-samples-CBC1_P1-CBC5_A1-chromosome-NC_040279.1
plot-samples-CBC1_P1-CBC5_A1-chromosome-NC_040280.1
plot-samples-CBC1_P1-CBC5_A1-chromosome-NC_040281.1
plot-samples-CBC1_P1-CBC5_A1-chromosome-NC_040282.1
plot-samples-CBC1_P1-CBC5_A1-chromosome-NC_040283.1
plot-samples-CBC1_P1-CBC5_A1-chromosome-NC_040284.1
CBC1_P1
CBC5_A1
0
1
2
3
4
5
6
Position (bp)
1e7
GT
0/0
0/1
1/1



PART 9: Contingency Table per Chromosome

In [64]:

```
import dataframe_image as dfi

for chromosome in chrom_len_00.index:
    chromosome_df = vcf_df_04[ vcf_df_04.CHROM == chromosome ]

    # reset chromosome_df indexes for contingency table
    chromosome_df.reset_index(inplace=True, drop=True)
    chromosome_ct = contingency_table(samples, chromosome_df, chromosome)
```

```
Contingency Table - Chromosome NC_040279.1

                CBC5_A1_GT
```

```
                        0/0   0/1  1/1 other
CBC1_P1_GT 0/0            0   610   16     0
           0/1         1719  3363  338     0
           1/1            7    45    0     0
           other          0     0    0     0

Contingency Table - Chromosome NC_040280.1

                 CBC5_A1_GT
                        0/0   0/1  1/1 other
CBC1_P1_GT 0/0            0   418  183     0
           0/1         1104  3096  297     0
           1/1          214    71    0     0
           other          0     0    0     0

Contingency Table - Chromosome NC_040281.1

                 CBC5_A1_GT
                        0/0   0/1  1/1 other
CBC1_P1_GT 0/0            0  1056  972     0
           0/1         1660  3848  172     0
           1/1          678   104    0     0
           other          0     0    0     0

Contingency Table - Chromosome NC_040282.1

                 CBC5_A1_GT
                        0/0   0/1    1/1 other
CBC1_P1_GT 0/0            0  1202   3620     0
           0/1         1729  4268  12853     0
           1/1          802   133      0     0
           other          0     0      0     0

Contingency Table - Chromosome NC_040283.1

                 CBC5_A1_GT
                        0/0   0/1   1/1 other
CBC1_P1_GT 0/0            0  1152  4315     0
           0/1         4164  3872  1155     0
           1/1         1281   160     0     0
           other          0     0     0     0

Contingency Table - Chromosome NC_040284.1

                 CBC5_A1_GT
                        0/0   0/1  1/1 other
CBC1_P1_GT 0/0            0   524  423     0
           0/1         1095  3145  259     0
           1/1          377    73    0     0
           other          0     0    0     0

Contingency Table - Chromosome NC_040285.1

                 CBC5_A1_GT
                        0/0   0/1   1/1 other
CBC1_P1_GT 0/0            0   590  1516     0
           0/1         3136  2158   814     0
           1/1         3442   176     0     0
           other          0     0     0     0

Contingency Table - Chromosome NC_040286.1

                 CBC5_A1_GT
                        0/0   0/1  1/1 other
CBC1_P1_GT 0/0            0   585  650     0
```

```
           0/1         1441  3551  169     0
           1/1         2062   202    0     0
           other          0     0    0     0

Contingency Table - Chromosome NC_040287.1

                 CBC5_A1_GT
                        0/0   0/1   1/1 other
CBC1_P1_GT 0/0            0   722  1699     0
           0/1         1624  2693  1056     0
           1/1         3263   178     0     0
           other          0     0     0     0

Contingency Table - Chromosome NC_040288.1

                 CBC5_A1_GT
                        0/0   0/1   1/1 other
CBC1_P1_GT 0/0            0   590   414     0
           0/1        16078  3873  2022     0
           1/1        11756   874     0     0
           other          0     0     0     0

Contingency Table - Chromosome NC_040289.1

                 CBC5_A1_GT
                        0/0   0/1  1/1 other
CBC1_P1_GT 0/0            0   693  132     0
           0/1         1824  4567  850     0
           1/1           25   129    0     0
           other          0     0    0     0
```

In [ ]: