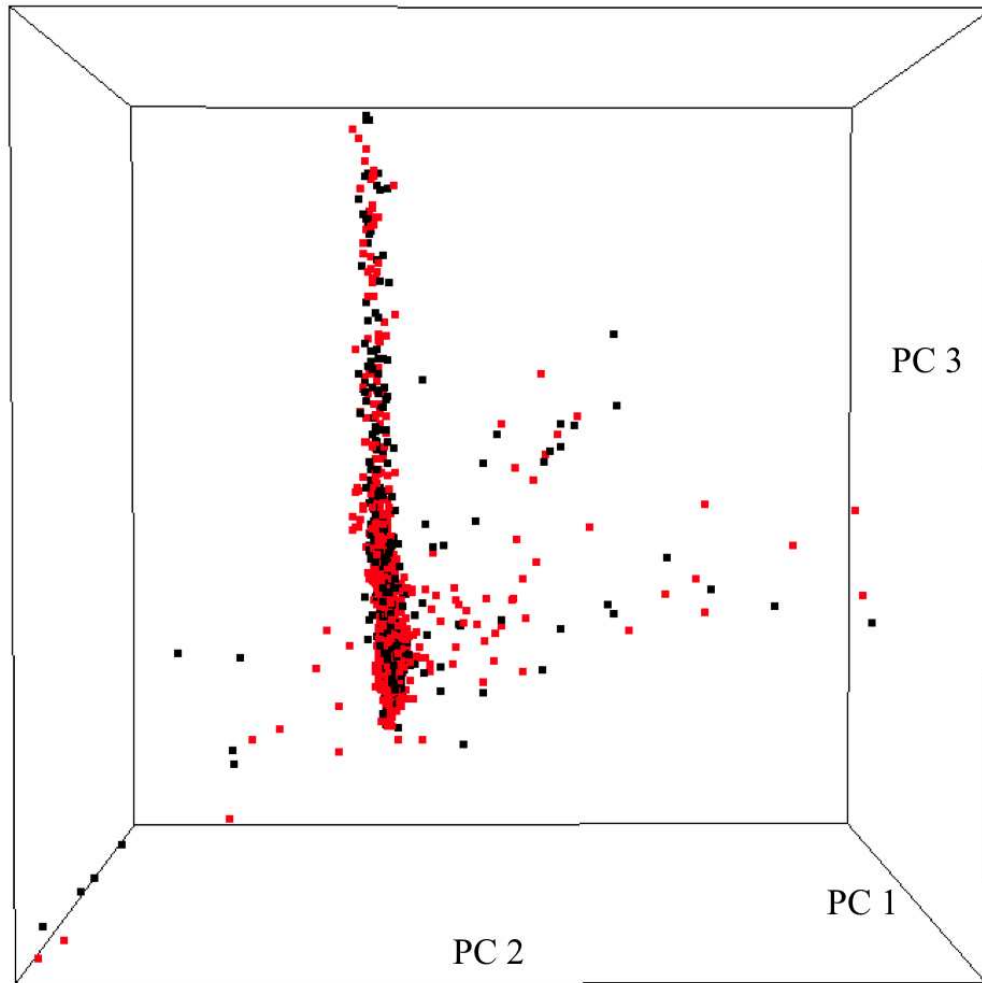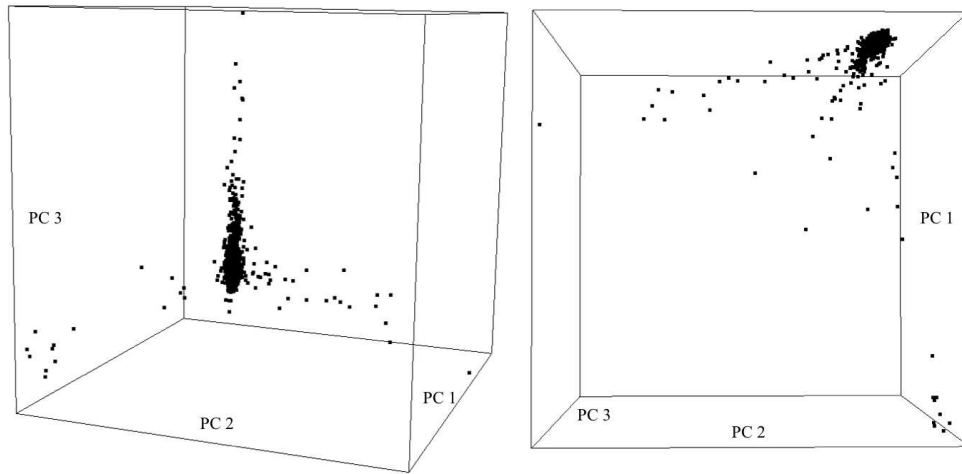Supplementary figure 1. Scree plots for the Canadian population study, the US extremes of phenotype study and the US sibling study.

Supplementary figure 2. Snapshot of the 3D plot of the top 3 principal components extracted from the Canadian population study sample alone. CF patients who are homozygous for F508del are colored in red; and those with other mutation types are colored in black. The red and black dots are evenly distributed alone each of the four arms, suggesting the population substructure observed in the Canadian population study samples is not due to difference in the CFTR mutation type.
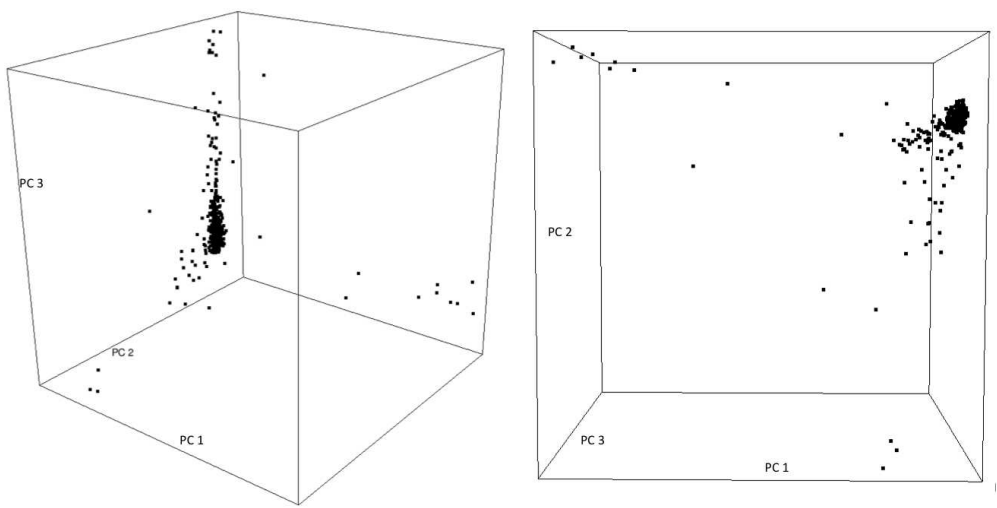
Supplementary figure 3. Two snapshots of the 3D plot of the top 3 principal components extracted from the US extremes of phenotype study samples. Three distinct arms are present, with most of the samples distributing along one arm.
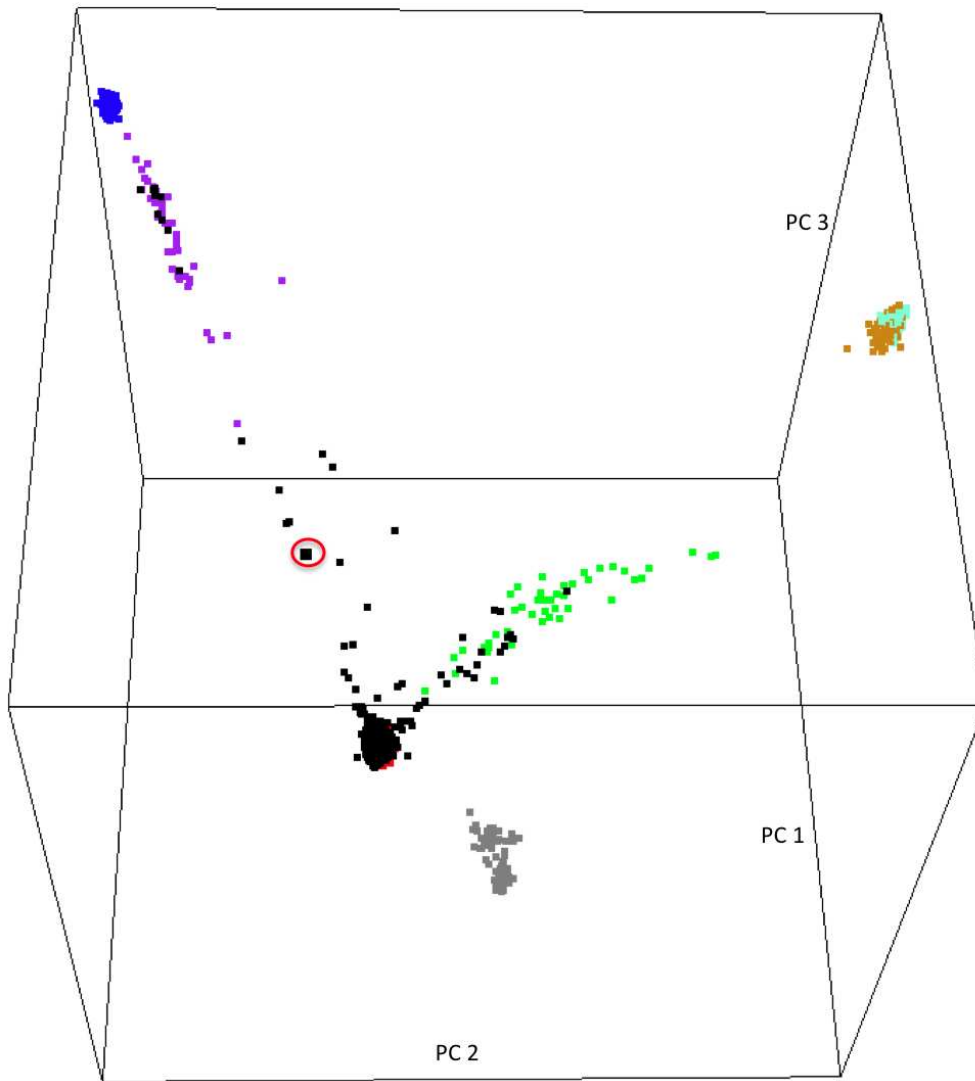
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
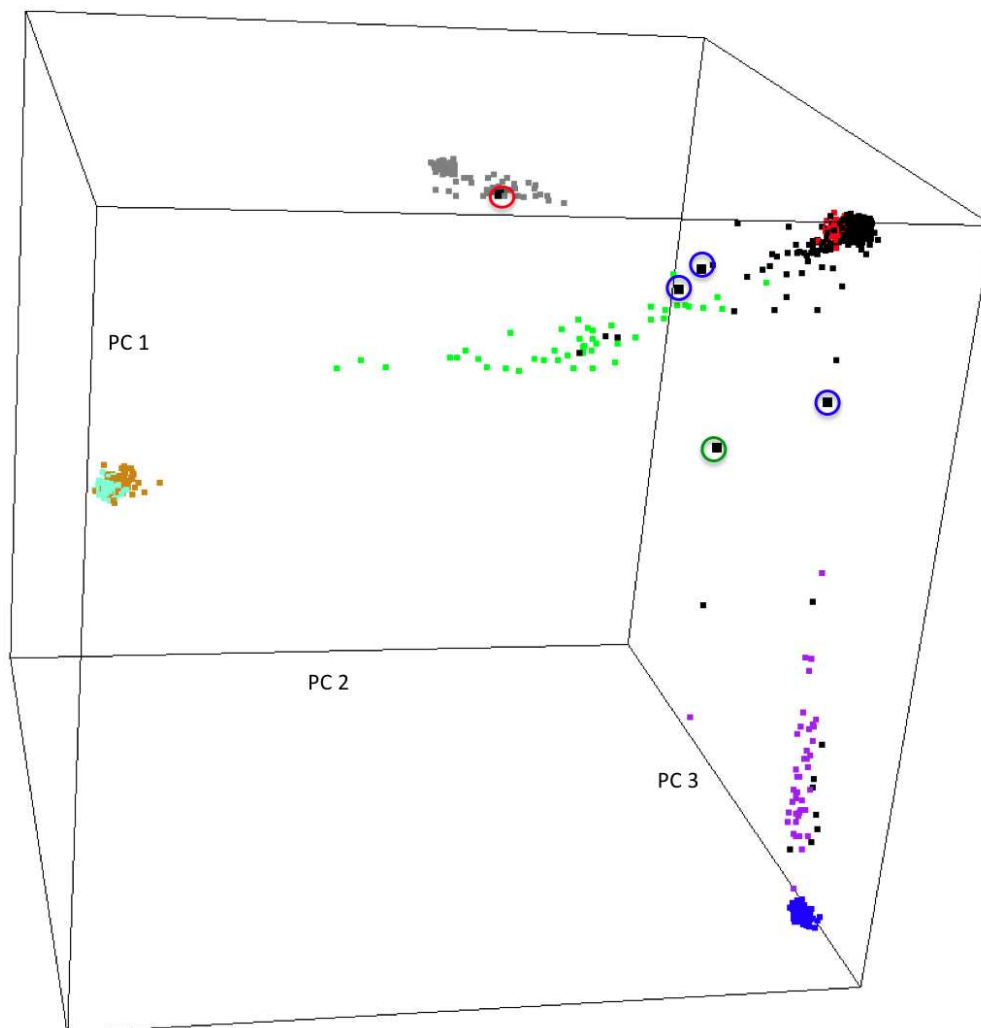52
53
54
55
56
57
58
59
60

Supplementary figure 4. Two snapshots of the 3D plot of the top 3 principal components extracted from the US sibling study samples. Three distinct arms are present, with most of the samples distributing along one arm.
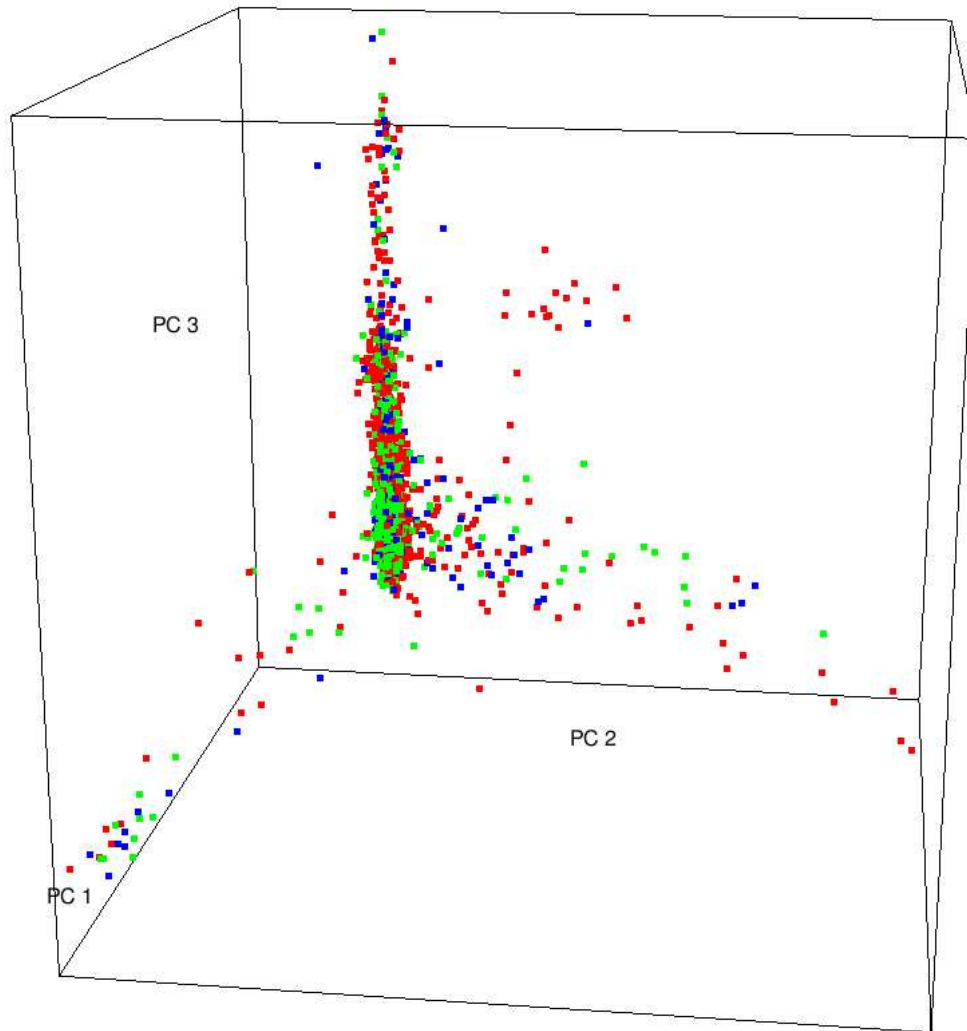
Supplementary figure 5. Snapshot of the 3D plot of the top 3 principal components extracted from the combined US extremes of phenotype study samples and the HapMap 3 samples. Color Legend: Black= US extremes of phenotype study sample, Purple=African American, Red= CEPH Caucasian, Orange=Han Chinese, Grey=Gujarati Indian, Aquamarine=Japanese, Green=Mexican, Blue=Yoruba. The point highlighted in the red circle indicates discrepancy between self-reported ethnicity and ethnicity estimated using one's genotype, where a self-reported 'Hispanic' sample distributed along the CEPH Caucasian-African American arm, suggesting his/her admixed Caucasian and African ancestries.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Supplementary figure 6. Snapshot of the 3D plot of the top 3 principal components extracted from the combined US sibling study samples and HapMap 3 samples. Color Legend: Black= US sibling study samples, Purple=African American, Red= CEPH Caucasian, Orange=Han Chinese, Grey=Gujarati Indian, Aquamarine=Japanese, Green=Mexican, Blue=Yoruba. Points highlighted in circles are indicative of discrepancy between self-reported ethnicity and ethnicity estimated using one's genotype. One self-reported 'Caucasian' subject (circled in green) distributed away from the CEPH Caucasian cluster; one self-reported 'Hispanic' individual (circled in red) scattered along the CEPH Caucasian-Gujarati Indian arm; and three individuals (circled in blue) with self-reported 'Asian' or 'Asian, Caucasian' ethnicity did not scatter around either of the CEPH Caucasian or the Asian clusters, one fell along the CEPH Caucasian-African American arm and the other two fell along the CEPH Caucasian-Mexican arm.

Supplementary figure 7. Snapshot of the 3D plot of the top 3 principal components extracted from the North American CF samples. Color legend: Red = Canadian population study samples, Green = US extremes of phenotype study samples, Blue = US sibling study samples. Individuals from each of three studies distributed evenly along 3 of the 4 arms; one of the arms consisted almost exclusively of individuals from the Canadian population study.
263x279mm (72 x 72 DPI)