Original Article

Enriching Targeted Sequencing Experiments for Rare Disease Alleles

Todd L. Edwards[1], Zhuo Song[2,3], and Chun Li[2,4,*]

[1]Vanderbilt Epidemiology Center, Division of Epidemiology, Vanderbilt University, Nashville, TN 37203

[2]Center for Human Genetics Research, [3]Department of Molecular Physiology and Biophysics, [4]Department of Biostatistics, Vanderbilt University, Nashville, TN 37212

Associate Editor:

SUPPLEMENTARY MATERIALS

Table 1. (CDRV) Ratio of average counts of disease alleles captured to the average maximum possible disease alleles (percent change relative to SampleSeq). K=0.01, assuming p_d=0.01 in calculation of E_{ik}

Size	Sample-Seq	Case Burden	Proxy Burden	Random Cases	Random Controls
50	0.31	0.31(-0.61)	0.19(-38.2)	0.15(-51.43)	0.04(-86.93)
100	0.33	0.33(0.52)	0.21(-35.34)	0.17(-47.78)	0.05(-85.95)
150	0.37	0.37(-0.11)	0.25(-32.41)	0.2(-45.43)	0.05(-85.32)
200	0.40	0.39(-1.33)	0.27(-32.96)	0.22(-43.67)	0.06(-84.84)
250	0.41	0.4(-1.98)	0.27(-34.4)	0.24(-42.16)	0.06(-84.44)
300	0.42	0.41(-2.56)	0.27(-34.99)	0.25(-40.86)	0.07(-84.09)
350	0.43	0.41(-2.93)	0.28(-34.99)	0.26(-39.69)	0.07(-83.77)
400	0.43	0.41(-3.46)	0.28(-34.77)	0.26(-38.53)	0.07(-83.46)
450	0.43	0.41(-3.65)	0.28(-34.25)	0.27(-37.44)	0.07(-83.16)
500	0.43	0.41(-4.06)	0.28(-34.03)	0.27(-36.68)	0.07(-82.96)

Table 2. (CDRV) Ratio of average counts of disease alleles captured to the average maximum possible disease alleles (percent change relative to SampleSeq). K=0.05, assuming p_d=0.01 in calculation of E_{ik}

Size	Sample-Seq	Case burden	Proxy burden	Random cases	Random controls
50	0.28	0.27(-2.7)	0.17(-41.18)	0.13(-53.6)	0.04(-87.09)
100	0.34	0.33(-3.35)	0.22(-34.89)	0.17(-49.94)	0.05(-86.07)
150	0.37	0.35(-4.96)	0.23(-36.96)	0.19(-48.25)	0.05(-85.6)
200	0.38	0.36(-4.6)	0.24(-37.13)	0.2(-45.65)	0.06(-84.87)
250	0.38	0.36(-3.9)	0.24(-37.21)	0.21(-43.84)	0.06(-84.37)
300	0.38	0.36(-4.06)	0.24(-35.75)	0.22(-42.16)	0.06(-83.9)
350	0.38	0.36(-4.71)	0.25(-34.44)	0.22(-41.07)	0.06(-83.6)
400	0.38	0.36(-4.44)	0.25(-34.11)	0.23(-39.82)	0.06(-83.25)
450	0.38	0.36(-5.33)	0.25(-34.21)	0.23(-38.95)	0.06(-83.01)
500	0.37	0.35(-5.03)	0.24(-34.27)	0.23(-37.72)	0.06(-82.67)

Table 3. (CDRV) Ratio of average counts of disease alleles captured to the average maximum possible disease alleles (percent change relative to SampleSeq). K=0.1, assuming p_d=0.01 in calculation of E_{ik}.

Size	Sample-Seq	Case Burden	Proxy Burden	Random Cases	Random Controls
50	0.24	0.24(-0.41)	0.14(-42.71)	0.13(-43.79)	0.03(-86.22)
100	0.27	0.26(-0.83)	0.16(-41.74)	0.16(-41.02)	0.04(-85.54)
150	0.27	0.27(-2.04)	0.16(-39.88)	0.17(-38.98)	0.04(-85.04)
200	0.28	0.27(-1.68)	0.16(-40.3)	0.17(-37.46)	0.04(-84.67)
250	0.28	0.27(-2.33)	0.16(-40.49)	0.18(-36.29)	0.04(-84.38)
300	0.28	0.27(-2.76)	0.16(-41.03)	0.18(-35.44)	0.04(-84.17)
350	0.28	0.27(-3.59)	0.16(-40.24)	0.18(-34.51)	0.04(-83.94)
400	0.27	0.26(-3.53)	0.16(-39.47)	0.18(-33.18)	0.04(-83.62)
450	0.27	0.26(-4.08)	0.17(-38.85)	0.18(-32.41)	0.05(-83.43)
500	0.29	0.28(-3.64)	0.18(-37.29)	0.2(-31.05)	0.05(-83.1)

Table 4. (CDRV) Ratio of average counts of disease alleles captured to the average maximum possible disease alleles (percent change relative to SampleSeq). K=0.2, assuming p_d=0.01 in calculation of E_{ik}.

Size	Sample-Seq	Case Burden	Proxy Burden	Random Cases	Random Controls
50	0.20	0.2(3.01)	0.11(-43.57)	0.13(-35.88)	0.03(-83.38)
100	0.22	0.22(0.76)	0.13(-43.18)	0.15(-35.05)	0.04(-83.16)
150	0.23	0.23(-0.81)	0.14(-40.73)	0.15(-33.51)	0.04(-82.76)
200	0.23	0.23(-0.16)	0.14(-40.1)	0.16(-31.34)	0.04(-82.2)
250	0.23	0.23(-0.92)	0.14(-38.93)	0.16(-30.15)	0.04(-81.89)
300	0.23	0.22(-2.06)	0.14(-39.29)	0.16(-29.22)	0.04(-81.65)
350	0.23	0.22(-2.56)	0.14(-39.07)	0.16(-28.4)	0.04(-81.44)
400	0.23	0.22(-2.53)	0.14(-38.59)	0.16(-27.55)	0.04(-81.22)
450	0.24	0.24(-2.09)	0.15(-38.03)	0.18(-26.61)	0.05(-80.98)
500	0.27	0.26(-2.01)	0.17(-37.42)	0.2(-25.69)	0.05(-80.74)

Table 5. (SA) Ratio of average counts of disease alleles captured to the average maximum possible disease alleles (percent change relative to SampleSeq). K=0.01, assuming p_d=0.01 in calculation of E_{ik}.

Size	Sample-Seq	Case Burden	Proxy Burden	Random Cases	Random Controls
50	0.66	0.66(-0.11)	0.54(-18.58)	0.63(-4.95)	0.14(-78.65)
100	0.69	0.69(-0.08)	0.54(-21.58)	0.65(-4.58)	0.15(-78.57)
150	0.72	0.72(-0.18)	0.58(-19.79)	0.69(-4.3)	0.15(-78.51)
200	0.74	0.74(-0.33)	0.59(-20.4)	0.71(-4.08)	0.16(-78.46)
250	0.76	0.75(-0.4)	0.58(-22.8)	0.73(-3.88)	0.16(-78.41)
300	0.77	0.76(-0.28)	0.57(-25.03)	0.74(-3.69)	0.17(-78.37)
350	0.77	0.77(-0.31)	0.58(-25.22)	0.75(-3.52)	0.17(-78.33)
400	0.78	0.78(-0.26)	0.59(-24.47)	0.76(-3.32)	0.17(-78.29)
450	0.80	0.8(-0.21)	0.61(-23.69)	0.78(-3.15)	0.17(-78.25)
500	0.82	0.81(-0.32)	0.62(-23.58)	0.79(-3.08)	0.18(-78.23)

Table 6. (CDCV) Ratio of average counts of disease alleles captured to the average maximum possible disease alleles (percent change relative to SampleSeq). K=0.01, assuming p_d=0.01 in calculation of E_{ik}.

Size	Sample-Seq	Case Burden	Proxy Burden	Random Cases	Random Controls
50	0.9	0.9(-0.13)	0.91(1.22)	0.48(-46.88)	0.4(-55.27)
100	0.91	0.92(0.26)	0.93(1.82)	0.51(-43.86)	0.43(-52.72)
150	0.91	0.91(-0.18)	0.94(2.45)	0.53(-41.74)	0.45(-50.94)
200	0.91	0.9(-0.92)	0.93(2.13)	0.54(-40.22)	0.46(-49.66)
250	0.9	0.89(-1.35)	0.92(1.57)	0.55(-39.06)	0.46(-48.68)
300	0.91	0.9(-1.66)	0.92(1.21)	0.57(-38.06)	0.48(-47.84)
350	0.92	0.9(-1.95)	0.93(1.28)	0.58(-37.04)	0.49(-46.98)
400	0.93	0.91(-2.48)	0.94(1.41)	0.59(-36.13)	0.5(-46.21)
450	0.93	0.9(-2.9)	0.94(1.62)	0.6(-35.17)	0.51(-45.4)
500	0.93	0.9(-3.15)	0.95(1.92)	0.61(-34.18)	0.51(-44.57)