Supplementary Table S1. Characteristics of the biomarker assays used for the FNIH OA Biomarkers Consortium sample measurements.

Biomarker	Biomarker Stability	US Status	Assay Type	Starting Dilution (volume required for duplicate analyses)	LLOQ*	Lowest reported value (number of extreme outliers)	Control Ranges
serum C1,2C	Unknown	RUO	C-I	1:2 (50 µl)	0.03 µg/ml	<0.06 µg/ml (0 (0))	LQC: 0.038-0.132 µg/mL HQC: 0.676-1.096 µg/mL
serum C2C-HUSA	Unknown	RUO	C-I	1:2 (50 µl)	10 ng/ml	<20 ng/mL (3 (1))	LQC: 12-46 ng/mL HQC: 213-400 ng/mL
serum Coll2-1NO2	Unknown	RUO	C-I	1:2 (50 µl)	0.469 nM	<0.9375 nM (5 (2))	LQC: 1.5-4.5 nM HQC: 10-20 nM
serum CPII	Maybe up to 1X freeze/thaw cycles	RUO	C-I	1:2 (50 µl)	50 ng/ml	<100 ng/mL (6 (3))	LQC: 170-334 ng/mL HQC: 942-1464 ng/mL
serum CS846	Unknown	RUO	C-I	1:5 (20 µl)	20 ng/ml	<100 ng/mL (8 (4))	LQC: 17-46 ng/mL HQC: 215-355 ng/mL
serum CTX-1	Up to 7X freeze/thaw cycles	IVD	C-I	Neat (100 µl)	0.149 ng/ml	<0.149 ng/mL (0 (0))	LQC: 0.244-0.366 ng/mL HQC: 0.787-1.181 ng/mL
serum COMP	Up to 5X; stable at least 7 days at 4°C	RUO	Sandwich ELISA	1:50 (4 µl)	4 ng/ml	<200 ng/mL (2 (2))	LQC: 10.4-15.6 ng/mL HQC: 37.0-55.6 ng/mL
serum HA	Up to 8X freeze/thaw cycles	RUO	Sandwich protein binding assay	1:11 (19-30 µl)	50 ng/ml	<50 ng/mL (7 (4))	LQC: 33-61 ng/mL MQC: 145-242 ng/mL HQC: 352-586 ng/mL
serum MMP-3	Unknown but analogous assay suggests up to at least 5 freeze thaw cycles	RUO	Sandwich ELISA	1:5 (20 µl)	0.62 ng/ml	<3.10 ng/mL (10 (7))	LQC: N/A ng/mL MQC: N/A ng/mL HQC: N/A ng/mL
serum NTXI	Up to 3X freeze/thaw cycles	IVD	C-I	1:5 (20 µl)	5 nM BCE	<5 nM BCE (2 (1))	LQC: 6.7-11.3 nm BCE HQC: 24.0-29.2 nm BCE
serum PIIANP	Manufacturer to provide data	RUO	C-I	1:2 (20 µl)	33 ng/ml	<66 ng/mL (0 (0))	LQC: 42-95 ng/mL HQC: 351-730 ng/mL
urine Col-2-1NO2	Manufacturer to provide data	RUO	C-I	1:2 (50 µl)	0.0625 nM	<0.0625 nM (8 (8))	LQC: 0.3-0.5 nM HQC: 1.5-2.5 nM

urine C1,2C	Unknown	RUO	C-I	1:2 (50 µl)	0.03 µg/ml	<0.06 µg/ml (1 (1))	LQC: 0.021-0.205 µg/mL HQC: 0.753-1.317 µg/mL
urine C2C-HUSA HUSA	Stable less than 24 hours at 4°C	RUO	Sandwich assay	Neat (100 µl)	156 pg/ml	<156 pg/mL (4 (4))	LQC: 223-463 pg/mL HQC: 2866-3973 pg/mL
urine CTXII	Up to 7X freeze/thaw cycles	RUO	C-I	Neat (80 µl)	0.59 ng/ml	<0.59 ng/mL (1 (1))	LQC: 1.13-1.70 ng/mL HQC: 3.19-4.78 ng/mL
urine NTXI	Up to 3X freeze/thaw cycles	IVD	C-I	Neat (100 µl)	20 nM BCE	<20 nM BCE (1 (1))	LQC: 350-464 nM BCE HQC: 1146-1518 nM BCE
urine CTXIα	Up to 7X freeze/thaw cycles	RUO	C-I	1:8 (12 µl)	0.30 ng/ml	<2.40 ng/mL (2 (2))	LQC: 12.54-18.82 ng/mL HQC: 21.25-31.87 ng/mL
urine CTX-1β	Up to 7X freeze/thaw cycles	IVD	C-I	1:4 (20 µl)	1.37 ng/ml	<5.48 ng/mL (1 (1))	LQC: 9.0-13.5 ng/mL HQC: 36.1-54.2 ng/mL
Urine Creatinine	Up to 4X freeze/thaw cycles	RUO	Colorimetric based on modified Jaffe method	1:40 (2.5 µl)	0	0 (0)	LQC: 6.0-7.9 mmol/L HQC: 21.6-28.6 mmol/L

RUO=research use only; IVD=in vitro diagnostic (for osteoporosis); C-I= competitive-inhibition ELISA;
LLOQ=lower limit of quantification represented by the lowest standard;
The lowest reported value is the lowest assay standard times the starting sample dilution factor;
An extreme outlier was defined as a value above 5 times the interquartile range above the third quartile;
Neat=undiluted; C-I=competitive inhibition ELISA;
*prior to creatinine normalization for urine markers

Supplementary Table S2. Non-transformed concentrations of biomarkers for each group of the FNIH OA Biomarkers Consortium project.

Biomarker (unit of measure)	**Cases (n=194)**	**Comparators (n=406)**		**JSL+Pain Progressors (n=194)**	**JSL only Progressors (n=103)**	**Pain only Progressors (n=103)**	**Non-Progressors (n=200)**
	Mean (SD) Median (range)	Mean (SD) Median (range)		Mean (SD) Median (range)	Mean (SD) Median (range)	Mean (SD) Median (range)	Mean (SD) Median (range)
Serum C1,2C (µg/ml)	0.39 (0.15) 0.38 (0.05-1.09)	0.38 (0.14) 0.37 (0.04-1.34)		0.39 (0.15) 0.38 (0.05-1.09)	0.38 (0.13) 0.37 (0.13-0.77)	0.38 (0.14) 0.38 (0.10-0.88)	0.38 (0.15) 0.37 (0.04-1.34)
Serum C2C-HUSA (ng/ml)	212.0 (54.9) 204.0 (102.0-423.0)	208.9 (46.8) 202.0 (100.0-395.0)		212.0 (54.9) 204.0 (102.0-423.0)	208.6 (45.4) 203.0 (114.0-395.0)	207.3 (45.6) 202.0 (115.0-339.0)	209.9 (48.2) 201.0 (100.0-336.0)
Serum COLL2-1 NO2 (nM)	8.91 (4.99) 8.20 (0.00-37.98)	8.93 (5.45) 7.97 (0.00-45.17)		8.91 (4.99) 8.20 (0.00-37.98)	8.26 (3.90) 7.61 (0.00-21.85)	9.40 (6.69) 7.97 (0.42-45.17)	9.03 (5.41) 8.20 (0.00-43.08)
Serum CPII (ng/ml)	944.2 (363.2) 894.0 (204.0-3006.0)	945.7 (391.0) 889.0 (252.0-4063.0)		944.2 (363.2) 894.0 (204.0-3006.0)	869.4 (280.9) 849.0 (354.0-2374.0)	1015.8 (473.8) 961.0 (252.0-3723.0)	948.9 (387.3) 887.0 (283.0-4063.0)
Serum CS846 (ng/ml)	79.2 (60.2) 65.0 (0.0-412.0)	76.8 (52.8) 66.0 (1.0-383.0)		79.2 (60.2) 65.0 (0.0-412.0)	67.8 (49.4) 55.0 (6.0-338.0)	79.5 (49.5) 73.5 (8.0-294.0)	80.1 (55.8) 69.0 (1.0-383.0)
Serum CTXI (ng/ml)	0.42 (0.21) 0.39 (0.07-1.23)	0.39 (0.22) 0.34 (0.08-1.79)		0.42 (0.21) 0.39 (0.07-1.23)	0.39 (0.20) 0.36 (0.09-1.10)	0.40 (0.26) 0.34 (0.11-1.79)	0.38 (0.21) 0.34 (0.08-1.33)
Serum COMP (ng/ml)	761.1 (280.1) 705.0 (168.0-1903.0)	783.4 (302.5) 739.0 (156.0-2007.0)		761.1 (280.1) 705.0 (168.0-1903.0)	777.4 (267.6) 743.5 (267.0-1712.0)	768.9 (290.5) 702.0 (219.0-1526.0)	794.0 (325.4) 749.0 (156.0-2007.0)
Serum HA (ng/ml)	49.5 (36.5) 38.0 (4.0-193.0)	45.2 (40.2) 34.0 (3.0-297.0)		49.5 (36.5) 38.0 (4.0-193.0)	49.2 (33.7) 44.0 (5.0-172.0)	47.2 (54.1) 28.5 (4.0-297.0)	42.1 (34.5) 33.0 (3.0-213.0)
Serum MMP-3	17.5 (11.3)	16.9 (11.2)		17.5 (11.3)	19.4 (13.0)	16.2 (10.1)	16.0 (10.7)

(ng/ml)	14.4 (1.0-76.3)	14.6 (0.0-99.9)		14.4 (1.0-76.3)	16.3 (3.7-99.9)	13.9 (3.5-57.2)	14.0 (0.0-88.8)
Serum NTXI (nm BCE)	15.5 (4.9) 15.0 (4.0-28.0)	14.8 (4.8) 14.0 (3.0-43.0)		15.5 (4.9) 15.0 (4.0-28.0)	14.8 (4.6) 14.0 (7.0-29.0)	14.9 (4.6) 15.0 (6.0-32.0)	14.8 (5.0) 14.0 (3.0-43.0)
Serum PIIANP (ng/ml)	2581.1 (783.7) 2524.0 (781.0-5219.0)	2677.2 (752.5) 2626.5 (160.0-5656.0)		2581.1 (783.7) 2524.0 (781.0-5219.0)	2550.0 (725.1) 2418.0 (780.0-4256.0)	2644.3 (710.8) 2656.0 (1035.0-4136.0)	2759.7 (779.9) 2759.5 (160.0-5656.0)
Urine Coll21NO2 creatinine adjusted (nM/mmol Cr)	0.025 (0.014) 0.021 (0.002-0.120)	0.024 (0.015) 0.020 (0.003-0.108)		0.025 (0.014) 0.021 (0.002-0.120)	0.023 (0.014) 0.021 (0.005-0.108)	0.025 (0.018) 0.020 (0.006-0.105)	0.024 (0.015) 0.019 (0.003-0.103)
Urine C1,2C creatinine adjusted (ng/mmol Cr)	0.010 (0.011) 0.007 (0.000-0.045)	0.011 (0.011) 0.009 (0.000-0.063)		0.010 (0.011) 0.007 (0.000-0.045)	0.012 (0.011) 0.010 (0.000-0.043)	0.013 (0.013) 0.011 (0.000-0.059)	0.010 (0.011) 0.007 (0.000-0.063)
Urine C2C-HUSA creatinine adjusted (ng/mmol Cr)	165.6 (103.0) 149.1 (0.0-695.1)	152.7 (90.3) 137.9 (0.0-763.3)		165.6 (103.0) 149.1 (0.0-695.1)	161.4 (79.4) 150.4 (0.0-511.0)	161.5 (112.6) 137.8 (0.0-763.3)	143.8 (82.1) 133.4 (0.0-435.6)
Urine CTXII creatinine adjusted (ng/mmol Cr)	333.09 (210.53) 283.65 (59.38-1446.38)	287.69 (190.75) 233.55 (0.00-1794.12)		333.09 (210.53) 283.65 (59.38-1446.38)	313.51 (228.62) 256.36 (58.33-1794.12)	312.07 (211.06) 261.11 (0.00-1240.00)	261.84 (152.00) 222.41 (0.00-1015.00)
Urine NTXI creatinine adjusted (nM BCE/mmol Cr)	34.9 (16.6) 32.1 (8.6-116.4)	32.6 (18.2) 29.1 (6.2-152.0)		34.9 (16.6) 32.1 (8.6-116.4)	31.3 (12.6) 29.4 (8.1-75.5)	34.8 (21.8) 31.1 (9.1-152.0)	32.1 (18.6) 28.4 (6.2-136.9)
Urine CTXIα creatinine adjusted (µg/mmol Cr)	0.46 (0.33) 0.39 (0.00-2.56)	0.41 (0.33) 0.34 (0.00-2.76)		0.46 (0.33) 0.39 (0.00-2.56)	0.39 (0.24) 0.35 (0.00-1.13)	0.43 (0.36) 0.34 (0.00-2.00)	0.41 (0.35) 0.32 (0.00-2.76)

Urine CTXIß creatinine adjusted (µg/mmol Cr)	2.34 (1.61) 2.04 (0.00-7.98)	2.19 (1.67) 1.79 (0.00-11.08)		2.34 (1.61) 2.04 (0.00-7.98)	2.09 (1.42) 1.82 (0.00-7.14)	2.29 (1.68) 1.89 (0.00-7.58)	2.19 (1.78) 1.76 (0.00-11.08)

JSL=joint space loss

Supplementary Table S3. Correlations between baseline biomarkers and covariates.

Biomarker (z scored)	Sex*	Pain Meds*	Race*	BL JSW+	BL WOMAC Pain+	BL Age+	BL BMI+
serum C1,2C	**-0.2381 (<.0001)**	0.0325 (0.4219)	**0.3425 (0.0404)**	0.0455 (0.2659)	-0.0347 (0.3960)	0.0552 (0.1773)	-0.0443 (0.2799)
serum C2C-HUSA	-0.1943 (0.4136)	**-0.1249 (0.0287)**	**0.4799 (0.0133)**	-0.0260 (0.5263)	0.0358 (0.3827)	0.0407 (0.3205)	0.0710 (0.0830)
serum COLL2-1 NO2	-0.4379 (0.1856)	-0.1313 (0.1098)	**0.5757 (<.0001)**	-0.0127 (0.7561)	0.0520 (0.2039)	0.0152 (0.7102)	-0.0622 (0.1288)
serum CPII	-0.2660 (0.5801)	**-0.0993 (<.0001)**	**0.8107 (<.0001)**	0.0113 (0.7834)	0.0219 (0.5925)	0.0304 (0.4584)	-0.0342 (0.4048)
serum CS846	**-0.1095 (0.0093)**	-0.0061 (0.6448)	0.0202 (0.6321)	-0.0056 (0.8922)	0.0243 (0.5539)	-0.0405 (0.3236)	0.0167 (0.6835)
serum CTXI	**-0.1479 (<.0001)**	0.1160 (0.0541)	**-0.1441 (0.0310)**	0.0267 (0.5145)	-0.0427 (0.2970)	-0.0152 (0.7103)	**-0.1228 (0.0026)**
serum COMP	0.1317 (0.1167)	0.0668 (0.7265)	0.0987 (0.0950)	0.0102 (0.8034)	-0.0627 (0.1260)	**0.2550 (<.0001)**	-0.0620 (0.1307)
serum HA	0.0010 (0.9821)	-0.0586 (0.4207)	**0.3268 (<.0001)**	**-0.1007 (0.0138)**	-0.0109 (0.7899)	**0.3829 (<.0001)**	0.0271 (0.5087)
serum MMP-3	**1.0617 (<.0001)**	-0.0067 (0.5986)	-0.1028 (0.1412)	0.0295 (0.4727)	-0.0684 (0.0953)	**0.0850 (0.0381)**	**-0.1057 (0.0098)**
serum NTXI	**-0.1662 (0.0001)**	0.1850 (0.5795)	0.0386 (0.1098)	0.0081 (0.8424)	-0.0054 (0.8961)	0.0538 (0.1886)	-0.0410 (0.3176)
serum PIIANP	-0.0377 (0.3946)	-0.0556 (0.5065)	0.3650 (0.5638)	-0.0100 (0.8066)	**0.0819 (0.0451)**	-0.0006 (0.9891)	**0.1517 (0.0002)**

urine Col2-1 NO2	**-0.2649 (0.0024)**	**-0.2227 (<.0001)**	**-0.2222 (<.0001)**	0.0167 (0.6836)	-0.0018 (0.9659)	0.0312 (0.4456)	-0.0246 (0.5486)
urine C1,2C	**0.2664 (0.0127)**	0.0304 (0.6111)	-0.0416 (0.4939)	0.0658 (0.1073)	0.0076 (0.8530)	**-0.1090 (0.0075)**	0.0761 (0.0628)
urine C2C-HUSA	**-0.1548 (0.0031)**	0.0322 (0.6740)	**0.0083 (0.0035)**	**-0.1006 (0.0138)**	**-0.0323 (0.4298)**	**0.2673 (<.0001)**	-0.0400 (0.3284)
urine CTXII	**-0.2952 (<.0001)**	0.0477 (0.3227)	0.1972 (0.1347)	-0.0621 (0.1287)	**0.0805 (0.0488)**	**0.1599 (<.0001)**	0.0584 (0.1533)
urine NTXI	**-0.4217 (<.0001)**	**0.1724 (0.0006)**	**-0.2061 (0.0029)**	-0.0245 (0.5497)	-0.0601 (0.1413)	0.0496 (0.2253)	**-0.1366 (0.0008)**
urine CTXIα	**-0.2870 (<.0001)**	**0.1704 (0.0249)**	-0.0413 (0.3738)	0.0082 (0.8421)	-0.0624 (0.1269)	0.0002 (0.9963)	**-0.0942 (0.0212)**
urine CTXIβ	**-0.3037 (<.0001)**	0.1817 (0.3298)	0.0263 (0.5589)	0.0282 (0.4904)	-0.0260 (0.5259)	-0.0599 (0.1431)	**-0.0887 (0.0300)**

BL=baseline; Urine biomarkers normalized to creatinine; JSW=joint space width;

*Difference in means (p-value) presented for dichotomous variables (presented as Sex: men - women; Medication use: no - yes; Race: nonwhite-white)

+Pearson correlation (p-value) presented for continuous variables

Values in bold represent associations that were significant at $P \leq 0.05$

Supplementary Table S4. Pearson correlations among baseline biomarker concentrations.

Biomarkers	sC1,2C	sC2C-HUSA	sColl2-1NO2	ssCPII	sCS846	sCTXI	sCOMP	sHA	sMMP3	sNTXI	sPIIANP	uColl2-1NO2/Cr	uC1,2C/Cr	uC2C-HUSA/Cr	uCTXII/Cr	uNTXI/Cr	uCTXIα/Cr	uCTXIβ/Cr
Serum C1,2C	1.0000	0.3930	0.3266	0.4854	0.0665	0.0231	0.0723	0.0539	-0.0753	-0.0504	0.0876	0.0495	-0.0596	0.0689	0.0941	0.0301	0.0427	0.0152
Serum C2C-HUSA	0.3930	1.0000	0.2673	0.4391	0.0192	-0.0128	0.1062	0.0874	-0.1296	-0.0647	0.1920	-0.0679	-0.0084	0.0685	0.1229	0.0141	0.0545	0.0154
Serum COLL2-1 NO2	0.3266	0.2673	1.0000	0.5059	0.0292	0.0026	0.0702	0.0236	-0.0778	0.0490	0.0339	-0.0055	0.0217	0.0533	0.0499	-0.0009	-0.0093	-0.0145
Serum CPII	0.4854	0.4391	0.5059	1.0000	-0.0208	0.0070	0.0590	0.0550	-0.0959	-0.0610	0.1390	-0.0574	-0.0622	0.0795	0.0925	0.0425	0.0527	0.0157
Serum CS846	0.0665	0.0192	0.0292	-0.0208	1.0000	0.0864	0.0149	-0.0051	-0.0418	0.1549	-0.0273	-0.0122	-0.0016	-0.0019	-0.0123	0.0748	0.0992	0.1125
Serum CTXI	0.0231	-0.0128	0.0026	0.0070	0.0864	1.0000	0.0175	0.0609	-0.0445	0.6558	-0.0106	0.0571	-0.0344	0.2649	0.3588	0.7750	0.7685	0.7611
Serum Comp	0.0723	0.1062	0.0702	0.0590	0.0149	0.0175	1.0000	0.2784	0.1560	0.0715	0.1018	-0.0082	0.0287	0.1154	0.1065	-0.0184	-0.0140	-0.0600
Serum HA	0.0539	0.0874	0.0236	0.0550	-0.0051	0.0609	0.2784	1.0000	0.1239	0.0996	0.0353	0.0774	-0.0178	0.2289	0.2505	0.0852	0.0510	0.0037
Serum MMP-3	-0.0753	-0.1296	-0.0778	-0.0959	-0.0418	-0.0445	0.1560	0.1239	1.0000	0.0154	-0.0247	-0.0412	0.0906	-0.0669	-0.1276	-0.1513	-0.1348	-0.1320
Serum NTXI	-0.0504	-0.0647	0.0490	-0.0610	0.1549	0.6558	0.0715	0.0996	0.0154	1.0000	-0.0088	-0.0110	-0.0877	0.1720	0.3264	0.5569	0.5586	0.5205
Serum PIIANP	0.0876	0.1920	0.0339	0.1390	-0.0273	-0.0106	0.1018	0.0353	-0.0247	-0.0088	1.0000	-0.0985	-0.0304	0.0175	0.0481	-0.0287	-0.0244	-0.0227
Urine Coll21NO2 creatinine adj	0.0495	-0.0679	-0.0055	-0.0574	-0.0122	0.0571	-0.0082	0.0774	-0.0412	-0.0110	-0.0985	1.0000	0.2490	0.0090	0.0783	0.2181	0.0609	0.1244
Urine C1,2C creatinine adj	-0.0596	-0.0084	0.0217	-0.0622	-0.0016	-0.0344	0.0287	-0.0178	0.0906	-0.0877	-0.0304	0.2490	1.0000	-0.0221	0.0481	0.0390	-0.0464	0.0861
Urine C2C-HUSA creatinine adj	0.0689	0.0685	0.0533	0.0795	-0.0019	0.2649	0.1154	0.2289	-0.0669	0.1720	0.0175	0.0090	-0.0221	1.0000	0.4970	0.3739	0.3459	0.2158
Urine CTXII creatinine adj	0.0941	0.1229	0.0499	0.0925	-0.0123	0.3588	0.1065	0.2505	-0.1276	0.3264	0.0481	0.0783	0.0481	0.4970	1.0000	0.4346	0.3808	0.3634
Urine NTXI creatinine adj	0.0301	0.0141	-0.0009	0.0425	0.0748	0.7750	-0.0184	0.0852	-0.1513	0.5569	-0.0287	0.2181	0.0390	0.3739	0.4346	1.0000	0.8907	0.8237

Urine CTXIα Cr adjusted	0.0427	0.0545	-0.0093	0.0527	0.0992	0.7685	-0.0140	0.0510	-0.1348	0.5586	-0.0244	0.0609	-0.0464	0.3459	0.3808	0.8907	1.0000	0.7760
Urine CTXIβ Cr adjusted	0.0152	0.0154	-0.0145	0.0157	0.1125	0.7611	-0.0600	0.0037	-0.1320	0.5205	-0.0227	0.1244	0.0861	0.2158	0.3634	0.8237	0.7760	1.0000

Supplementary Table S5. Performance of 24M TIC continuous biomarkers (10-fold Cross Validated).

Mechanism	Biomarker	p-value (unadj)	Adjusted			AUC			IDI	NRI
			OR	OR 95% CI	p-value	base model*	new model	differ -ence		
Bone Resorption	Serum CTXI	0.0097	1.28	1.08, 1.53	0.0051	0.550	0.576	0.026	0.0103	0.2406
Inflammation	Serum HA	0.0281	1.22	1.01, 1.48	0.0415	0.551	0.566	0.015	0.0052	0.2060
Bone Resorption	Serum NTXI	0.0202	1.25	1.05, 1.48	0.0131	0.551	0.572	0.021	0.0074	0.1081
Bone Resorption	Urine CTX-1a	0.0052	1.32	1.11, 1.58	0.002	0.550	0.586	0.036	0.0129	0.2446
Bone Resorption	Urine CTX-1ß	0.0298	1.27	1.06, 1.51	0.0086	0.550	0.573	0.024	0.0094	0.1029
Cartilage Degradation	Urine C2C-HUSA	0.0078	1.27	1.06, 1.53	0.0108	0.550	0.568	0.019	0.0077	0.1639
Cartilage Degradation	Urinary CTXII	0.0016	1.37	1.15, 1.65	0.0006	0.550	0.587	0.037	0.0152	0.2487
Bone Resorption	Urine NTXI	0.017	1.29	1.08, 1.54	0.0057	0.550	0.580	0.030	0.0101	0.1226

AUC=area under curve; NRI=net reclassification index; IDI=integrated discrimination improvement; TIC=time integrated concentration
Urine markers are creatinine adjusted.
* Base model includes age, sex, race, BMI, baseline joint space width, baseline (WOMAC) pain, baseline Kellgren-Lawrence grade, pain medication use

Supplementary Table S6. Cut-offs used for creating categorical variables in combinatorial models to predict case status (based on 24M TICs for final models shown in Table 5).

Biomarker	n	Z-score related			Raw concentration related*			
		cut-off	min	max	min	max	mean	std
serum CTXI 24M TIC	447	z <= 0.5	-1.7	0.5	0.15	0.97	0.62	0.20
	151	z > 0.5	0.5	4.0	0.97	2.26	1.29	0.31
serum HA 24M TIC	221	z <= -0.5	-1.3	-0.5	5	62	41.9	13.0
	298	-0.5 < z <= 1	-0.5	1.0	62.5	164.5	102.1	28.6
	76	z > 1	1.0	6.3	166	526	234.6	71.6
serum NTXI 24M TIC	511	z <= 1	-2.4	0.9	7.5	38.5	27.1	6.1
	86	z > 1	1.0	5.9	39	84	46.1	7.3
urine CTXIα 24M TIC creatinine adj	212	z <= -0.5	-1.4	-0.5	0.00	0.53	0.33	0.14
	253	-0.5 < z <= 0.5	-0.5	0.5	0.54	1.12	0.81	0.17
	133	z > 0.5	0.5	7.4	1.13	5.10	1.65	0.60
urine C2C-HUSA 24M TIC creatinine adj 24M	197	z <= -0.5	-1.9	-0.5	0.00	235.74	164.14	51.44
	318	-0.5 < z <= 1	-0.5	1.0	236.88	479.97	335.95	67.63
	83	z > 1	1.0	7.7	483.07	1574.52	615.03	157.44
urine CTXII 24M TIC creatinine adj 24M	78	z <= -1	-1.7	-1.0	48.9011	279.420	220.10	50.98
	151	-1 < z <= -0.5	-1.0	-0.5	282.893	449.236	369.17	47.95
	370	z > -0.5	-0.5	5.1	451.199	2352.91	807.56	303.22
urine NTXI 24M TIC creatinine adj 24M	206	z <= -0.5	-1.8	-0.5	12.62	50.92	39.06	8.49
	312	-0.5 < z <= 1	-0.5	1.0	51.05	95.41	69.09	11.82
	81	z > 1	1.0	6.4	96.07	254.29	121.61	28.38

* serum CTXI, HA and PIIANP in ng/ml; serum NTXI in nm BCE; urine CTXIα and CTXIβ in µg/mmol Cr; urine C2C-HUSA and CTIXII in ng/mmol Cr; urine NTXI in nM BCE/mmol Cr

Supplementary Table S7. Hierarchical models for urine biomarkers (based on 24M TICs).

	Model 1:	Model 2:	Model 3:	Model 4:	Model 5:
	UR-CTXII only	UR-CTXII, UR-ALPHA	UR-CTXII, UR-ALPHA, UR-NTXI	UR-CTXII, UR-ALPHA, UR-NTXI, UR-C2C-HUSA	UR-CTXII, UR-ALPHA, UR-C2C-HUSA
C-statistic (AUC)*	0.583	0.609	0.611	0.618	0.617
UR-CTXII	p=0.0006	p=0.0126	p=0.0149	p=0.0930	p=0.0884
z < -1 vs. -1 ≤ z < -0.5	0.5 (0.2, 0.9)	0.5 (0.2, 1.0)	0.5 (0.2, 1.0)	0.5 (0.3, 1.1)	0.5 (0.3, 1.1)
z ≥ -0.5 vs. -1 ≤ z < -0.5	1.5 (1.0, 2.3)	1.3 (0.9, 2.1)	1.3 (0.9, 2.1)	1.2 (0.8, 1.9)	1.2 (0.8, 1.9)
UR-ALPHA		p=0.1108	p=0.4129	p=0.4165	p=0.1331
z < -0.5 vs. -0.5 ≤ z < 0.5		0.8 (0.5, 1.2)	0.8 (0.5, 1.4)	0.8 (0.5, 1.3)	0.8 (0.5, 1.2)
z ≥ 0.5 vs. -0.5 ≤ z < 0.5		1.3 (0.9, 2.0)	1.3 (0.7, 2.3)	1.3 (0.7, 2.2)	1.3 (0.8, 2.0)
UR-NTXI			p=0.9802	p=0.9898	
z < -0.5 vs. -0.5 ≤ z < 1			1.0 (0.6, 1.6)	1.0 (0.6, 1.6)	
z ≥ 1 vs. -0.5 ≤ z < 1			1.0 (0.5, 1.9)	1.0 (0.5, 1.9)	
UR-C2C-HUSA				p=0.5628	p=0.5575
z < -0.5 vs. -0.5 ≤ z < 1				0.8 (0.5, 1.3)	0.8 (0.5, 1.3)
z ≥ 1 vs. -0.5 ≤ z < 1				1.1 (0.7, 1.9)	1.1 (0.7, 1.8)

C-Statistics (AUCs) not cross validated; UR=urine; None of the urine markers are statistically significant when added in addition to Urine-CTXII.

Supplementary Table S8. Hierarchical models for baseline biomarkers predicting case status at 48M.

	Model 1:	Model 2:	Model 3:	Model 4:	Model 4:	Model 6:
	UR-CTXII only	UR-CTXII, UR-ALPHA	UR-CTXII, UR-NTXI	UR-CTXII, SER-NTXI	UR-CTXII, SER-CTXI,	UR-CTXII, UR-ALPHA, SER-NTXI
C-statistic (AUC)*	0.563	0.577	0.573	0.586	0.576	0.588
UR-CTXII	p=0.0065	p=0.0185	p=0.0144	p=0.0136	p=0.0145	p=0.0226
z < -1 vs. -1 ≤ z < -0.5	0.4 (0.2, 1.0)	0.4 (0.2, 1.0)	0.4 (0.2, 1.0)	0.4 (0.2, 1.0)	0.4 (0.2, 1.0)	0.4 (0.2, 1.0)
z ≥ -0.5 vs. -1 ≤ z < -0.5	1.4 (0.9, 2.1)	1.3 (0.9, 2.0)	1.3 (0.9, 2.0)	1.4 (0.9, 2.1)	1.4 (0.9, 2.1)	1.3 (0.9, 2.0)
UR-ALPHA		p=0.1443				p=0.4012
z ≥ 0.5 vs. z < 0.5		1.4 (0.9, 2.1)				1.2 (0.8, 1.9)
UR-NTXI			p=0.3713			
z ≥ 0.5 vs. z < 0.5			1.2 (0.8, 1.8)			
SER-NTXI				p=0.0394		p=0.0889
z ≥ 1 vs. z < 1				1.6 (1.0, 2.5)		1.5 (0.9, 2.4)
SER-CTXI					p=0.4667	
z < -0.5 vs. -0.5 ≤ z < 0.5					1.0 (0.7, 1.5)	
z ≥ 0.5 vs. -0.5 ≤ z < 0.5					1.3 (0.8, 2.0)	

* C-Statistics (AUCs) not cross validated; ur=urine; ser=serum

Supplementary Text for "Predictive Validity of Biochemical Biomarkers in Knee Osteoarthritis – Data from the FNIH Biomarkers Consortium" (Kraus et al.)

PATIENTS AND METHODS

Study Design

In brief, eligible participants for the present study were those with at least one knee with a Kellgren-Lawrence grade (KLG) of 1-3 at baseline from central reading and availability at baseline and 24M of medial joint space width data from knee radiographs, knee magnetic resonance images (MRI), stored serum and urine specimens and clinical data. Participants with knee or hip replacement between baseline and 24M were excluded to avoid potential effects of this surgery on systemic biomarker levels. Participants were also excluded if they had radiographic and pain progression by 12M follow-up. Knees were excluded that had lateral joint space narrowing (JSN) grade 2 or 3 at baseline.

Definitions of radiographic and symptomatic progression

Radiographic progression was defined by loss of minimum JSW in the medial compartment of ≥0.7 mm from baseline to 24, 36 or 48M from radiographs obtained by non-fluoroscopic fixed flexion protocol (SynaFlexor, Synarc, Newark, CA)[1]. Knee pain was assessed using the Western Ontario McMasters (WOMAC) pain subscale[2]. Based upon an established minimum clinically important difference (MCID) for pain worsening[3], persistent pain progression was defined as a pain increase of ≥9 points at 2 or more timepoints (on a 0-100 normalized score, 100=worst) from the 24M to 60M pain assessment.

For better covariate balance among the groups, the knees selected for the four groups were frequency matched to the extent feasible, using KLG strata 1-3 and body mass index (BMI) strata <25, 25-27.5, 27.5-30, 30-35 and ≥35 kg/m^2.

Biospecimen Collection

Serum (s) and urine (u, unspun) were obtained at baseline, 12M and 24M. By design, all subjects had baseline and 24M samples. A total of 15 subjects lacked 12M data and samples. Overall a total of N=1,785 specimens from 600 subjects were available for analysis. The majority of subjects provided fasting blood and urine samples (defined as more than 8 hours without food prior to specimen collection); the proportions of fasting samples at baseline, 12M and 24M were 98%, 96% and 98%, respectively for serum, and 92%, 93% and 94%, respectively for urine. Biospecimens were provided from the OAI sample repository by Fisher Scientific. Encoded and unthawed stock samples of serum (0.5 ml x 2 aliquots) and urine (2 ml aliquot) were provided to LabCorp Clinical Trials (San Leandro, CA) for these analyses. The stock urine sample was aliquoted by LabCorp and an aliquot provided to Artialis (Liege, Belgium) for their analysis of urine Col2-1 NO2. All biospecimens were encoded by personnel at the study's data coordinating center at UCSF (the data and sample supervising institution). The two analytic sites were provided knowledge of the clustering of samples by individual in order to run all samples for a particular individual on the same assay plate and thereby minimize within subject technical variability. Both sites were blinded to the timing (baseline, 12M, 24M) of specimens. The biomarker data were forwarded to UCSF for unblinding and linking to the clinical data for subsequent analysis by personnel at the central statistical analysis center (Brigham and Women's Hospital).

Biomarker Assays

As part of the recommendations to advance the science of biomarkers, the OARSI / FDA Biomarkers Working Group[4] recommended measurement of a broad set of biomarkers in the same sample set. By consensus of the 20 authors, the following commercially available biomarkers, were recommended for inclusion in a future study to provide comparative data and biological insights from which to continue to assess the utility and relevance of an array of established OA-related biomarkers: urinary CTX-II, serum COMP, serum hyaluronan, serum and urine C1, 2C, serum and urine C2C, serum and urine Coll2-1 and Coll2-1NO2, serum CPII, Serum PIIANP, urine/serum NTX-1, urine/ serum CTX-1, serum CS846, and serum MMP-3. This panel was considered an initial starting point for a process of OA biomarker qualification in evolution. In addition, to be selected for study, the biomarker had to be available "off the shelf" as a commercially available kit, and available world-wide. Each of these biomarkers met this criterion and therefore this entire list was chosen for the initial phase I analysis of the FNIH OA Biomarker Consortium study. This does not imply that they necessarily represent the sole or best possibilities for fulfilling the needs of the OA drug development process. Formal biomarker qualification is acknowledged to be an arduous and dynamic process, benefitting from an advancing base of knowledge to expand and refine the contexts for which biomarkers are formally qualified. However, this study is important for establishing a paradigm by which OA-related biomarker qualification can proceed. The sequencing of assays was deliberately designed and standardized after discussion with each kit manufacturer, to minimize freeze thaws and prioritize analyses of known labile biomarkers and those for which freeze-thaw stability was unknown (Supplementary Table 1). Great care was taken to use kits with the same lot number for all sample analyses of each particular biomarker. Inter-assay coefficients of variation (CVs, provided in Table 1) are based on the average CVs of high and low concentration control samples provided with each kit and run on each plate; when available (as for serum MMP-3 and HA), a third medium concentration standard was also run on each plate and combined with the high and low concentrations for an overall CV. All samples were run in duplicate. The initial dilutions for sample measurements (listed in Supplementary Table 1) were agreed upon by the kit manufacturers. Samples with concentrations above the highest standard

were repeated at a higher dilution until results were within the linear range of the assay; thus all high values were quantifiable. When biomarker results were below the lowest standard, the kit manufacturer was consulted to determine lower dilutions that could be tested without the likelihood of incurring problems with assay inhibition. For many of the biomarkers, there were still appreciable numbers of samples with values below the LLOQ. Several imputation strategies were considered. For the purposes of these analyses, concentrations below the lower limits of detection were imputed by interpolation from the standard curve extended from the lowest standard to zero. This was deemed superior to random imputation, particularly for biomarkers, such as HA and CS846 whose standard curves were clearly linear below the lowest standard (data not shown). For Col2-1 NO2, concentrations below LLOQ were imputed as 80% of the lowest standard. Both of these methods are variations on the single imputation method for dealing with values below the lower limit of detection[5]. For each biomarker and visit we identified extreme outliers. Typically, an outlier is defined as an observation more than 1.5 times the interquartile range (IQR) above the third quartile or above the first quartile[6]. To identify extreme outliers, we identified those observations that were >5 times the IQR above the third quartile, or >3 times the IQR above the third quartile when the biomarker was transposed to the log scale. These observations were excluded from the main analysis. As a sensitivity analysis, we imputed the extreme outliers with the maximum value of the non-excluded observations for each biomarker/time point.

Statistical analysis

We conducted the analysis in several steps. First, we evaluated each biomarker separately, using a logistic regression model, after adjusting for age, sex, BMI, baseline radiographic joint space narrowing, baseline WOMAC pain and baseline use of pain medications. For each biomarker, we assessed the p-value (those with $p<0.1$ were advanced to multivariable modeling) and discriminative ability using the c-statistic (AUC), category-less net reclassification index (NRI), and the integrated discrimination improvement (IDI) index. The additional discriminative indices were used for confirmatory purposes and the

p-value served as a main determinant of significance. Ten-fold cross validation was used to assess the prediction error; this process was repeated 100 times to generate a range for each cross-validated measure of discrimination. We also examined cross correlation among biochemical markers to avoid issues of collinearity.

Second, for each biomarker selected on the basis of p-value, we created a 5-level categorical variable to evaluate the dose-response relationship between the biomarker and risk of case status. Categories were based on z-score-based deviation from the mean. To improve the statistical power, categories were combined based on intermediary analysis if adjacent categories exhibited similar relationship with the outcome.

Third, we further reduced the data by selecting among the correlated markers ($r \geq 0.80$), those that showed the best discriminative ability, based on the combination of c-statistics (AUCs), NRI, IDI. We evaluated both the continuous and categorical biomarkers in this step. We conducted a number of sensitivity analyses to confirm that the selection process did not affect the performance of the final model.

Finally, we conducted multivariable analyses. Based on univariable results from the primary analyses, we selected biomarkers to advance to multivariable modeling. We built models in a hierarchical fashion, with the best performing biomarkers added to the model first (performance based on p-value, OR, IDI, AUC, NRI as described above). The added predictive ability of each new biomarker was assessed by the p-value and OR of the newly added marker in the multivariable model. The biomarkers with adjusted p-values>0.1 were eliminated from the final models. All multivariable models were adjusted for age, sex, BMI, baseline radiographic joint space narrowing, baseline WOMAC pain and baseline use of pain medications. To improve the transparency of results and interpretation, we present three sets of best models: urine biomarkers only, serum biomarkers only and models based on the combination of serum

and urine biomarkers. We repeated the ten-fold cross validation process for the final models. Multivariable modeling was done only for the primary analysis of pain and radiographic joint space loss progression versus comparator knees lacking the combination of pain and radiographic progression and only for 24M TIC and baseline biomarker concentrations.

Sensitivity Analysis for Outliers

In total, out of 33,915 measurements, we excluded 61 extreme outliers for 35 participants. The results of the analysis remained largely the same when we imputed the maximum of the non-extreme outlier values for each extreme outlier. The main difference in the sensitivity analysis was that the 24 month TIC for serum-HA no longer met the $p<0.10$ threshold in unadjusted analysis ($p=0.14$). There were three controls with extreme values of serum-HA that were excluded from the main analysis; including these participants narrowed the difference between the cases and controls (z-scores for controls vs. cases of -0.06 vs. 0.13 in original analysis and -0.04 vs. 0.09 in sensitivity analysis) while simultaneously increasing the variability around the serum HA measurement.

Biomarker characteristics

The biomarker concentrations for cases and comparators in the primary analysis (two groups) are provided in Table 1. Non-transformed concentrations based on further stratification by four groups (three progressor and one non-progressor) are provided in Supplementary Table 2. Associations between each biomarker under consideration and set of person-based baseline characteristics (age, sex, BMI, baseline radiographic joint space narrowing, baseline WOMAC pain and baseline use of pain medications) are presented in Supplementary Table 3. Correlations among all the biomarkers are presented in

Supplementary Table 4. The only collinear biomarker (Pearson r>0.8) was uNTXI that correlated with both uCTXIα and uCTXIß. There were also strong correlations (r>0.75) between sCTXI and uNTXI, uCTXIα and uCTXIß and between uCTXIα and uCTXIß. The strongest correlation for uCTXII was with the other collagen type II degradation marker, uC2C (r=0.50); uCTXII was also modestly correlated with the collagen type I degradation biomarkers (r=0.33-0.43).

REFERENCES

1. Peterfy C, Li J, Saim S, et al. Comparison of fixed-flexion positioning with fluoroscopic semi-flexed positioning for quantifying radiographic joint-space width in the knee: test-retest reproducibility. *Skeletal Radiol* 2003;32:128-132.

2. Bellamy N, Buchanan WW, Goldsmith CH, et al. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 1988;15(12):1833-1840.

3. Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. *Arthritis Rheum* 2001;45(4):384-391.

4. Kraus VB, Burnett B, Coindreau J, et al. Application of biomarkers in the development of drugs intended for the treatment of osteoarthritis. *Osteoarthritis Cartilage* 2011;19(5):515-542.

5. Vexler A, Tao G, Chen X. A toolkit for clinical statisticians to fix problems based on biomarker measurements subject to instrumental limitations: from repeated measurement techniques to a hybrid pooled-unpooled design. *Methods Mol Biol* 2015;1208:439-460.

6. Moore D, McCabe G. Introduction to the Practice of Statistics. 3rd ed. New York: W. H. Freeman, 1999.