Transcriptomic landscape of breast cancers through mRNA sequencing

Jeyanthy Eswaran[1, 2], Dinesh Cyanam[1], Prakriti Mudvari[1,2] , Sirigiri Divijendra Natha Reddy[2], Suresh B Pakala[2], Sujit S Nair[2], Liliana Florea[5] Suzanne A. W. Fuqua[3] Sucheta Godbole and Rakesh Kumar[1, 2,*]

[1]McCormick Genomic and Proteomics Center, [2]Department of Biochemistry and Molecular Biology, The George Washington University, Washington, DC 20037, USA [4]Breast Center, Baylor College of Medicine, One Baylor Plaza, 1220 Alkek, Houston, TX 77030, USA. [5]McKusick-Nathans Institute of Genetic Medicine, School of Medicine, Johns Hopkins University, Baltimore 21205

[*]Correspondence: bcmrxk@gwumc.edu.

Supplement Table 1 : mRNA sequencing of all breast cancer samples and their alignment statistics overview against Ensembl GRCh37.62 B. reference using TopHat aligning program

Group	Sample	No. of Reads	No. of filtered Reads	Total aligned Reads	No. of unique aligned reads
TNBC	A1	60,710,416	60,685,374	44318616	9,042,922
	A2	45604510	45,599,673	27176808	5,717,561
	A3	70,754,133	70,722,004	49593712	10,632,860
	A4	19,578,597	58,642,366	37267560	7,713,545
	A5	54,369,476	54,358,396	33873616	7,232,310
	A6	55834686	55,820,801	34714784	7,365,416
Non-TNBC	B1	91,454,047	91,436,856	66784320	13,227,246
	B2	67,655,694	67,636,543	47344552	8,062,318
	B3	73,057,434	73,035,978	50850020	8,897,604
	B4	88,014,336	87,997,294	64734596	12,727,038
	B5	77,050,214	77,031,995	52807956	11,378,537
	B6	67,991,540	67,982,906	49895320	8,957,757
HER2-Positive	C1	107,397,991	107,376,005	83112352	16,184,345
	C2	76,188,935	76,173,402	58157296	10,183,562
	C3	54,369,476	86,518,204	64519232	7,232,310
	C4	90,948,038	90,930,481	67870968	12,765,356
	C5	30,277,252	30,273,376	22253992	3,778,143

Supplement Table 2: Alignment statistics summary of all 17 breast cancer sample mRNA paired - end reads aligning to human genome Ensembl GRCh37.62 B. using TopHat aligning program

Group	Sample	No. of Reads	Aligned reads	Mate Missing	Both Mate Flip	One Mate Flip
TNBC	A1	60,685,374	44318616 (73%)	14616456 (24%)	1089002 (2%)	661298 (1%)
	A2	45,599,673	27176808 (60%)	16496051 (36%)	914022 (2%)	1012792 (2%)
	A3	70,722,004	49593712 (70%)	18853234 (27%)	1518464 (2%)	756596 (1%)
	A4	58,642,366	37267560 (64%)	19006216 (32%)	1603415 (3%)	765174 (1%)
	A5	54,358,396	33873616 (62%)	18457006 (34%)	1118918 (2%)	908856 (2%)
	A6	55,820,801	34714784 (62%)	18727324 (34%)	1598970 (3%)	779772 (1%)
Non-TNBC	B1	91,436,856	66784320 (73%)	21429496 (23%)	2226656 (2%)	996382 (1%)
	B2	67,636,543	47344552 (70%)	18059232 (27%)	1572350 (2%)	660410 (1%)
	B3	73,035,978	50850020 (70%)	19728358 (27%)	1778810 (2%)	678790 (1%)
	B4	87,997,294	64734596 (74%)	21010828 (24%)	1940694 (2%)	311176 (0%)
	B5	77,031,995	52807956 (69%)	21464000 (28%)	1957316 (3%)	804722 (1%)
	B6	67,982,906	49895320 (73%)	16418898 (24%)	1446812 (2%)	221872 (0%)
HER2-Positive	C1	107,376,005	83112352 (77%)	20711532 (19%)	3099878 (3%)	452244 (0%)
	C2	76,173,402	58157296 (76%)	15891740 (21%)	1856422 (2%)	267940 (0%)
	C3	86,518,204	64519232 (75%)	18548268 (21%)	3087692 (4%)	363010 (0%)
	C4	90,930,481	67870968 (75%)	19376264 (21%)	3399408 (4%)	283844 (0%)
	C5	30,273,376	22253992 (74%)	7007024 (23%)	912776 (3%)	99584 (0%)

Supplement Table 3: Read distribution summary showing where the aligned fragments map in human genome (Ensembl GRCh37.62 B.)

Group	Sample	Exons	Introns	Intergenic	Junctions
TNBC	A1	20492832	19959500	5617556	1434864
	A2	17356158	12182789	581590	12182789
	A3	23494250	17860829	10516303	949758
	A4	22756230	16460393	420267	1554030
	A5	20338611	14226532	936903	1408172
	A6	20820289	14245836	2028149	1470783
Non-TNBC	B1	42017270	22853064	5138262	3205088
	B2	25899214	20024335	3655221	1457139
	B3	28159959	21918952	3229893	1519656
	B4	32924717	29752607	4310404	1689159
	B5	27741918	21256627	3750244	1807929
	B6	34603957	34558818	2566045	1250095
HER2-Positive	C1	42529319	42364696	2393313	1269324
	C2	29872983	27764770	1771141	1877579
	C3	36597306	28614421	2644367	1445994
	C4	36597306	28614421	2758681	1206588
	C5	13728557	8469355	1068602	804307

Supplement Table 4: Summary of total number of exons, unique exons and transcripts in each sample derived from cufflinks assembler using **A.** Ensembl GRCh37.62 **B.** hg19 UCSC database

A

Groups	Sample	Unique Exon	Total Exons	Transcripts	Genes
TNBC	A1	716839	321539	84060	17649
	A2	656272	288266	74702	14950
	A3	699294	317925	80919	18290
	A4	694691	306919	81063	15983
	A5	704720	310450	81607	16222
	A6	681530	303291	78652	16119
Non-TNBC	B1	715570	321884	85784	16927
	B2	704781	313280	82424	16536
	B3	710214	317292	83440	16738
	B4	710414	315059	82310	16690
	B5	717966	321313	84562	17038
	B6	677214	298393	77423	15488
HER2-Positiv	C1	679650	304608	80737	15969
	C2	679026	300684	78453	15588
	C3	664540	294511	76028	15322
	C4	695274	308421	80675	16005
	C5	653898	282700	71913	14648

B

Group	Sample	Uniq Exons	Total Exons	Transcripts
TNBC	A1	221390	541232	52859
	A2	198068	497645	46854
	A3	224390	537227	52535
	A4	205338	503748	48070
	A5	202729	500348	47288
	A6	206060	501733	47868
Non-TNBC	B1	215203	522329	50699
	B2	210217	509489	48738
	B3	212798	514139	49265
	B4	221390	541232	52859
	B5	197775	500814	46645
	B6	215374	529514	50794
HER2-Positive	C1	204993	498805	48172
	C2	205010	507543	48171
	C3	202729	500348	47288
	C4	210066	517095	49146
	C5	197775	500814	46645

Supplement Table 5: The number of low (less than 1FPKM) and high (higher than 10FPKM) abundant transcripts as well as the total transcripts in each sample

Groups	Sample	FPKM Bins Sum < 1	FPKM Bins Sum > 10	Total Transcripts
TNBC	A1	30357	1682	84060
	A2	27809	1695	74702
	A3	32718	1072	80919
	A4	29967	1889	81063
	A5	30190	1866	81607
	A6	29630	1837	78652
Non-TNBC	B1	31078	2718	85784
	B2	30541	1639	82424
	B3	31462	1602	83440
	B4	33895	940	82310
	B5	32208	1781	84562
	B6	30174	1271	77423
HER2-positive	C1	32762	791	80737
	C2	30422	1071	78453
	C3	31007	877	76028
	C4	31886	1042	80675
	C5	25212	1842	71913

Supplement Table 6: Spearman Correlation coefficient of samples based on **A.** gene expression and **B.** Transcript expression in all three groups

A

	A1	A2	A3	A4	A5	A6	B1	B2	B3	B4	B5	B6	C1	C2	C3	C4	C5
A1	1	0.87581495	0.84580995	0.87722101	0.89228084	0.89539837	0.86336786	0.84083686	0.85160928	0.82669325	0.84199441	0.81801669	0.84577362	0.86078516	0.85717914	0.82529943	0.83915184
A2	0.87581495	1	0.84790882	0.88860495	0.90576319	0.88769479	0.84883352	0.83961738	0.8437133	0.84228234	0.8400857	0.80345409	0.82341552	0.82953797	0.82575579	0.81599948	0.85273936
A3	0.84580995	0.84790882	1	0.85347546	0.85828002	0.84759419	0.85522917	0.84750292	0.86839949	0.86697511	0.85801779	0.82309945	0.8601228	0.79738037	0.85263336	0.84736033	0.86030732
A4	0.87722101	0.88860495	0.85347546	1	0.90409093	0.89307829	0.86063108	0.83931558	0.85532384	0.83939435	0.84656279	0.80590217	0.83332102	0.83258441	0.83302458	0.81586678	0.85052524
A5	0.89228084	0.90576319	0.85828002	0.90409093	1	0.89791797	0.87524738	0.8556358	0.87045706	0.86162926	0.86544489	0.81799633	0.85000335	0.83870471	0.8487111	0.83184121	0.86894388
A6	0.89539837	0.88769479	0.84759419	0.89307829	0.89791797	1	0.85399708	0.82479321	0.83914283	0.82388806	0.83338861	0.80166261	0.82762487	0.84837773	0.84267663	0.81437577	0.84785515
B1	0.86336786	0.84883352	0.85522917	0.86063108	0.87524738	0.85399708	1	0.88753421	0.89927185	0.86736602	0.89760482	0.86328544	0.84153387	0.80936511	0.85425824	0.83335448	0.86428318
B2	0.84083686	0.83961738	0.84750292	0.83931558	0.8556358	0.82479321	0.88753421	1	0.94043242	0.87459642	0.90748765	0.87712056	0.83950598	0.80407648	0.8399397	0.83790925	0.86092111
B3	0.85160928	0.8437133	0.86839949	0.85532384	0.87045706	0.83914283	0.89927185	0.94043242	1	0.8907349	0.92914532	0.88036983	0.86313361	0.81259747	0.85723982	0.85199703	0.87511963
B4	0.82669325	0.84228234	0.86697511	0.83939435	0.86162926	0.82388806	0.86736602	0.87459642	0.8907349	1	0.8783913	0.84671091	0.86114426	0.79223611	0.8463195	0.85636599	0.8748681
B5	0.84199441	0.8400857	0.85801779	0.84656279	0.86544489	0.83338861	0.89760482	0.90748765	0.92914532	0.8783913	1	0.88221316	0.84132556	0.79325651	0.84155794	0.83747637	0.86508113
B6	0.81801669	0.80345409	0.82309945	0.80590217	0.81799633	0.80166261	0.86328544	0.87712056	0.88036983	0.84671091	0.88221316	1	0.83011785	0.78783987	0.83468077	0.82573832	0.83991437
C1	0.84577362	0.82341552	0.8601228	0.83332102	0.85000335	0.82762487	0.84153387	0.83950598	0.86313361	0.86114426	0.84132556	0.83011785	1	0.8333189	0.89540517	0.87862711	0.8562385
C2	0.86078516	0.82953797	0.79738037	0.83258441	0.83870471	0.84837773	0.80936511	0.80407648	0.81259747	0.79223611	0.79325651	0.78783987	0.8333189	1	0.84424187	0.8159075	0.80242458
C3	0.85717914	0.82575579	0.85263336	0.83302458	0.8487111	0.84267663	0.85425824	0.8399397	0.85723982	0.8463195	0.84155794	0.83468077	0.89540517	0.84424187	1	0.8740703	0.86537531
C4	0.82529943	0.81599948	0.84736033	0.81586678	0.83184121	0.81437577	0.83335448	0.83790925	0.85199703	0.85636599	0.83747637	0.82573832	0.87862711	0.8159075	0.8740703	1	0.84760375
C5	0.83915184	0.85273936	0.86030732	0.85052524	0.86894388	0.84785515	0.86428318	0.86092111	0.87511963	0.8748681	0.86508113	0.83991437	0.8562385	0.80242458	0.86537531	0.84760375	1

B

	A1	A2	A3	A4	A5	A6	B1	B2	B3	B4	B5	B6	C1	C2	C3	C4	C5
A1	1	0.6920937	0.68075316	0.70825257	0.72121784	0.7312966	0.71227691	0.68066585	0.69065701	0.65178146	0.69098879	0.6454114	0.65053717	0.66857475	0.67272699	0.63216381	0.65536595
A2	0.6920937	1	0.65805478	0.73564509	0.7471427	0.70380477	0.70210404	0.69148137	0.69586135	0.65525115	0.69556395	0.62877982	0.61869159	0.62661019	0.63432784	0.61562996	0.65808681
A3	0.68075316	0.65805478	1	0.66648036	0.66944902	0.68115204	0.67730926	0.6700699	0.69150549	0.67960817	0.68629205	0.64038098	0.66066147	0.61435368	0.65854039	0.64458529	0.65924974
A4	0.70825257	0.73564509	0.66648036	1	0.75401281	0.72095712	0.72612975	0.69511989	0.70446611	0.66299577	0.70943327	0.63136093	0.63828502	0.64100646	0.64851956	0.62344349	0.66493747
A5	0.72121784	0.7471427	0.66944902	0.75401281	1	0.72059161	0.73386962	0.70660322	0.72003279	0.67889373	0.72255802	0.64451527	0.64652519	0.64110428	0.65493605	0.6318421	0.68158892
A6	0.7312966	0.70380477	0.68115204	0.72095712	0.72059161	1	0.70035569	0.66244473	0.67684202	0.64812958	0.68093762	0.62836823	0.63012231	0.6479903	0.65180498	0.62034423	0.66407799
B1	0.71227691	0.70210404	0.67730926	0.72612975	0.73386962	0.70035569	1	0.74025237	0.75302782	0.68585039	0.76087106	0.69026124	0.65173031	0.63831542	0.67288781	0.64427871	0.69130039
B2	0.68066585	0.69148137	0.6700699	0.69511989	0.70660322	0.66244473	0.74025237	1	0.7906518	0.69355617	0.76243351	0.69873372	0.6494727	0.62915597	0.65761847	0.64981198	0.68124671
B3	0.69065701	0.69586135	0.69150549	0.70446611	0.72003279	0.67684202	0.75302782	0.7906518	1	0.70800093	0.79004624	0.70091309	0.66573268	0.63531418	0.66960948	0.66021628	0.69401598
B4	0.65178146	0.65525115	0.67960817	0.66299577	0.67889373	0.64812958	0.68585039	0.69355617	0.70800093	1	0.70001041	0.68970594	0.69556764	0.64475294	0.68954028	0.69442598	0.7036168
B5	0.69098879	0.69556395	0.68629205	0.70943327	0.72255802	0.68093762	0.76087106	0.76243351	0.79004624	0.70001041	1	0.7059017	0.65105709	0.61757751	0.65592586	0.64877225	0.69084752
B6	0.6454114	0.62877982	0.64038098	0.63136093	0.64451527	0.62836823	0.69026124	0.69873372	0.70091309	0.68970594	0.7059017	1	0.6680185	0.6409413	0.68016487	0.6632465	0.67626726
C1	0.65053717	0.61869159	0.66066147	0.63828502	0.64652519	0.63012231	0.65173031	0.6494727	0.66573268	0.69556764	0.65105709	0.6680185	1	0.68912106	0.73886142	0.71958596	0.66850272
C2	0.66857475	0.62661019	0.61435368	0.64100646	0.64110428	0.6479903	0.63831542	0.62915597	0.63531418	0.64475294	0.61757751	0.6409413	0.68912106	1	0.6969237	0.66747065	0.63556426
C3	0.67272699	0.63432784	0.65854039	0.64851956	0.65493605	0.65180498	0.67288781	0.65761847	0.66960948	0.68954028	0.65592586	0.68016487	0.73886142	0.6969237	1	0.71485656	0.685548
C4	0.63216381	0.61562996	0.64458529	0.62344349	0.6318421	0.62034423	0.64427871	0.64981198	0.66021628	0.69442598	0.64877225	0.6632465	0.71958596	0.66747065	0.71485656	1	0.66166842
C5	0.65536595	0.65808681	0.65924974	0.66493747	0.68158892	0.66407799	0.69130039	0.68124671	0.69401598	0.7036168	0.69084752	0.67626726	0.66850272	0.63556426	0.685548	0.66166842	1

Supplement Table 7: Pair wise differentially expressing transcripts matrix generated from cuffdiff statistical analysis between each sample. The transcripts used for the analysis are reconstructed using **A.** Ensembl GRCh37.62 **B.** hg19 UCSC database

A

	A1	A2	A3	A4	A5	A6	**A**	B1	B2	B3	B4	B5	B6	**B**	C1	C2	C3	C4	C5	**C**
A1		**4624**	**4437**	**4212**	**3962**	**3938**	**835**	**5449**	**4456**	**4131**	**4553**	**4557**	**4774**		**4696**	**4389**	**5009**	**4890**	**4604**	
A2	4624		4937	2242	2217	2501	447	5017	3599	3709	3892	3827	3431		4565	3949	2139	3544	4121	
A3	4437	4937		4726	4449	4296	1200	6125	4506	4160	4090	4336	4720		5604	4624	4847	4559	4336	
A4	4212	2242	4726		2359	2648	522	4445	3641	3652	3951	3677	3788		4578	4022	2997	3895	4161	
A5	3962	2217	4449	2359		2632	447	4350	3429	3348	3712	3506	3551		4546	3898	3915	3810	4005	
A6	3962	2501	4296	2648	2632		530	4877	3968	3792	4117	3924	3791		4400	3922	3052	3840	4125	
A	3962	447	1200	522	447	530														
B1	3962	5017	6125	4445	4350	4877			4372	4222	5328	4111	5082	1317	6128	5404	5167	5687	5705	
B2	3962	3599	4506	3641	3429	3968		4372		2189	3380	2854	3180	531	4654	3841	3711	3630	3802	
B3	3962	3709	4160	3652	3348	3792		4222	2189		3149	2399	3104	455	4447	3567	3664	3455	3552	
B4	3962	3892	4090	3951	3712	4117		5328	3380	3149		3706	3278	839	4383	3348	3668	3089	3087	
B5	3962	3827	4336	3677	3506	3924		4111	2854	2399	3706		3414	563	5054	4213	4046	4156	4173	
B6	3962	3431	4720	3788	3551	3791		5082	3180	3104	3278	3414		771	4388	3408	2909	3162	3500	
B								1317	531	455	839	563	771							
C1	4696	4565	5604	4578	4546	4400		6128	4654	4447	4383	5054	4388			3523	4865	3873	3476	1069
C2	4389	3949	4624	4022	3898	3922		5404	3841	3567	3348	4213	3408		3523		3551	2735	2503	567
C3	5009	2139	4847	2997	3915	3052		5167	3711	3664	3668	4046	2909		4865	3551		3249	3843	1326
C4	4890	3544	4559	3895	3810	3840		5687	3630	3455	3089	4156	3162		3873	2735	3249		2441	590
C5	4604	4121	4336	4161	4005	4125		5705	3802	3552	3087	4173	3500		3476	2503	3843	2441		621
C															1069	567	1326	590	621	

B

	A1	A2	A3	A4	A5	A6	**A**	B1	B2	B3	B4	B5	B6	**B**	C1	C2	C3	C4	C5	**C**
A1		4624	4437	4212	3962	3938	835	5449	4456	4131	4553	4557	4774		4696	4389	5009	4890	4604	
A2	4624		4937	2242	2217	2501	447	5017	3599	3709	3892	3827	3431		4565	3949	2139	3544	4121	
A3	4437	4937		4726	4449	4296	1200	6125	4506	4160	4090	4336	4720		5604	4624	4847	4559	4336	
A4	4212	2242	4726		2359	2648	522	4445	3641	3652	3951	3677	3788		4578	4022	2997	3895	4161	
A5	3962	2217	4449	2359		2632	447	4350	3429	3348	3712	3506	3551		4546	3898	3915	3810	4005	
A6	3938	2501	4296	2648	2632		530	4877	3968	3792	4117	3924	3791		4400	3922	3052	3840	4125	
A	835	447	1200	522	447	530														
B1	5449	5017	6125	4445	4350	4877			4372	4222	5328	4111	5082	1317	6128	5404	5167	5687	5705	
B2	4456	3599	4506	3641	3429	3968		4372		2189	3380	2854	3180	531	4654	3841	3711	3630	3802	
B3	4131	3709	4160	3652	3348	3792		4222	2189		3149	2399	3104	455	4447	3567	3664	3455	3552	
B4	4553	3892	4090	3951	3712	4117		5328	3380	3149		3706	3278	839	4383	3348	3668	3089	3087	
B5	4557	3827	4336	3677	3506	3924		4111	2854	2399	3706		3414	563	5054	4213	4046	4156	4173	
B6	4774	3431	4720	3788	3551	3791		5082	3180	3104	3278	3414		771	4388	3408	2909	3162	3500	
B								1317	531	455	839	563	771							
C1	4696	4565	5604	4578	4546	4400		6128	4654	4447	4383	5054	4388			3523	4865	3873	3476	1069
C2	4389	3949	4624	4022	3898	3922		5404	3841	3567	3348	4213	3408		3523		3551	2735	2503	567
C3	5009	2139	4847	2997	3915	3052		5167	3711	3664	3668	4046	2909		4865	3551		3249	3843	1326
C4	4890	3544	4559	3895	3810	3840		5687	3630	3455	3089	4156	3162		3873	2735	3249		2441	590
C5	4604	4121	4336	4161	4005	4125		5705	3802	3552	3087	4173	3500		3476	2503	3843	2441		621
C															1069	567	1326	590	621	

Supplement Table 8: (A)The exons identified from intronic region and the number of transcripts that contain them (B) The table showing how many transcripts contain how many “exons from introns”.

A

Group	Sample	Total transcripts that comprise "exon from intron"	Unique transcripts that comprise "exon from intron" event	Common exons from intron event	Common exons from intron event
TNBC	A1	191153	26689	285	78
	A2	26438	6870		
	A3	234030	32580		
	A4	39079	9826		
	A5	83672	13229		
	A6	64458	13110		
Non-TNBC	B1	88865	15243	317	
	B2	97652	16356		
	B3	101187	17049		
	B4	102636	15749		
	B5	89480	17080		
	B6	53946	10316		
HER2-Positive	C1	158371	17196	244	
	C2	111892	14753		
	C3	83672	13229		
	C4	128482	15711		
	C5	13269	4439		

B

Group	**Sample**	**Number of Exons**				
		One	**Two**	**Three**	**Four**	**Five and above**
TNBC	A1	189952	468	46	14	12
	A2	26050	168	12	3	1
	A3	232652	530	59	22	10
	A4	38558	226	13	3	3
	A5	50481	232	20	8	4
	A6	63875	250	21	4	1
Non-TNBC	B1	88087	304	32	10	5
	B2	96690	373	43	17	4
	B3	100275	343	51	7	7
	B4	101873	293	27	5	12
	B5	88626	342	45	6	1
	B6	53404	209	24	8	4
HER2 Positive	C1	157419	366	33	18	8
	C2	111209	276	23	7	5
	C3	82974	266	37	8	4
	C4	127711	306	28	11	6
	C5	12992	117	13		1

Supplement Table 9: (A) The number of un annotated transcripts identified in each sample.
(B) The table showing how many un annotated transcripts contain how many exons.

A

Group	Sample	Previously unannotated Exons	Total number of transcripts that are unannotated	Total number of primary transcripts associated with unannotated transcript
TNBC	A1	55138	1021	871
	A2	5145	393	353
	A3	148326	1081	936
	A4	7754	553	488
	A5	8992	601	531
	A6	10500	582	536
Non-TNBC	B1	15894	855	757
	B2	17976	776	670
	B3	19061	838	720
	B4	19115	840	710
	B5	20668	927	814
	B6	10684	393	544
HER2-Positive	C1	23736	816	705
	C2	15307	626	539
	C3	16398	681	575
	C4	18506	735	639
	C5	3988	338	303

B

Group	**Sample**	**No.of Exons**								
		Two	**Three**	**Four**	**Five**	**Six**	**Seven**	**Eight**	**Nine**	**Ten+**
TNBC	A1	733	149	66	39	16	7	5	5	2
	A2	232	63	47	33	9	2	5	3	0
	A3	826	166	47	24	10	6	1	2	0
	A4	358	97	43	22	15	8	5	2	4
	A5	384	113	52	32	7	9	5	0	0
	A6	385	111	43	23	9	6	4	0	2
Non-TNBC	B1	553	151	64	38	16	11	10	3	10
	B2	504	149	62	27	15	9	2	0	9
	B3	562	145	57	33	15	8	5	4	10
	B4	554	139	60	43	13	15	6	5	6
	B5	648	152	52	41	12	7	7	1	8
	B6	389	113	54	34	17	12	8	2	10
HER2 Positive	C1	486	132	74	37	29	19	13	9	18
	C2	383	104	51	42	21	14	6	1	5
	C3	419	114	67	38	13	13	7	0	11
	C4	449	141	68	25	17	17	11	2	5
	C5	185	70	41	22	9	5	3	0	4

Supplement Table 10: Microarray analysis data collection

GEO Datasets used for Microarray Analysis:

GPL96

	No. of Samples		
GEO ID	**non-TNBC**	**TNBC**	**HER-2 Positive**
GSE7390	8	39	12
GSE2603	4	32	3
GSE3494	4	30	21
GSE2990	2	24	13
GSE2034	13	56	19
GSE11121	4	22	10
GSE1561	2	18	4
GSE20194	5	68	25
Total	42	289	107

GPL570

	No. of Samples		
GEO ID	**non-TNBC**	**TNBC**	**HER-2 Positive**
GSE7904	0	21	6
GSE2109	5	94	28
GSE19615	3	36	10
GSE12276	8	64	22
Total	16	215	66

Supplement Table 11: Differentially expressing genes from the microarray analysis is compared against the RNA sequencing comparison

GLP750					
Comparison	**RNA-Seq**	**MicroArray**	**Agree**	**Disagree**	**%Agree**
Non-TNBC Vs HER2-positive	2616	385	113	0	100
TNBC Vs HER2-positive	3088	1323	167	11	93.82022472
TNBC Vs Non-TNBC	2971	2033	562	2	99.64539007
GLP96					
Comparison	**RNA-Seq**	**MicroArray**	**Agree**	**Disagree**	**%Agree**
Non-TNBC Vs HER2-positive	2616	231	32	4	88.88888889
TNBC Vs HER2-positive	3088	593	144	11	92.90322581
TNBC Vs Non-TNBC	2971	734	316	1	99.68454259

Supplement Table 12: Gene signature comparison of this mRNA sequencing study compared against the published microarray based gene signatures.

A

Hu Classifier (Overlapping gene names are shown in Table 12B)				
Breast cancer Comparison	**Genes found in RNA-Seq**	**Microarray Intrinsic Gene List**	**Genes found in both both studies**	**Genes following the same expression trend**
Non-TNBC Vs HER2-positive	2972	306	73	58
TNBC Vs HER2-positive	3089	306	76	50
TNBC Vs Non-TNBC	2617	306	77	73

B

Comparison of the gene signature found in Hu et al against mRNA sequencing findings		
Non-TNBC vs HER-2 positive comparison	**TNBC vs HER-2 positive comparison**	**TNBC Vs Non-TNBC comparison**
AHCYL1	AK2	AHCYL1
AMFR	BUB1	AMFR
ATP6AP1	CASP7	ASF1A
BAG4	CBR4	BLVRA
BLVRA	CD59	BUB3
BUB3	CDC123	C10orf116
C10orf116	CDH3	C4A
CCNH	CHI3L1	CBR4
CD38	CHI3L2	CCNG2
CENPF	CIR1	CCNH
CES2	CLTC	CDH3
CLGN	CP	CENPF
CX3CL1	CRABP1	CHI3L1
DNAJC12	CSRP2	CHI3L2
DUSP6	DUSP4	CKS1B
ERBB2	DUSP6	CLGN
ESR1	ERBB2	CLTC
FUT8	FUBP1	CRABP1
GART	GRB7	CSRP2
GATA3	GSTK1	CX3CL1
GRB7	H3F3B	DNAJC12
GSTK1	HTATIP2	DUSP4
GSTM3	ID4	EIF2C2
HTATIP2	IFI30	ESR1
IFI6	IFI6	FABP7
IGFBP2	IGBP1	FOXC1
IRF9	IMPA2	FUBP1
ISG15	KDM5A	FUCA1
MMP7	KRT18	FUT8
MYO6	MFAP2	GATA3
NDUFB5	MXI1	GATM
NINJ1	NMI	GSTM3
PCM1	PAPSS2	ID4
PDCD4	PDCD5	IGBP1
PDCD5	PDHA1	IL13RA1
PDXDC1	PDIA4	IMPA2
PHYH	PDXDC1	KIAA0020
PIGT	PISD	KRT18
POLB	PLOD1	LFNG
PRR13	PRKACB	MFAP2
PSMC6	PRKCI	MMP1
PTGER3	RABEP1	MMP7
RCAN1	RBX1	MXI1
RNF10	RCC1	NFIB
RRAGA	SKP1	NINJ1
SLC39A6	SOX4	NMI
SLC9A3R1	STAT1	NRIP1
SOD2	TFF3	PCM1
SRPK1	WDR26	PDCD4
STAT1	WNK1	PDIA4
SYNGR1		PIGT
TCEAL1		PISD
TOP2B		PLOD1
TSG101		PPP2R5A
WNK1		PRKACB
YIF1A		PRNP
ZBTB4		PRR13
ZNF24		PTGER3
		RFC4
		RNF10
		SDC1
		SFRP1
		SKP1
		SLC39A6
		SLC9A3R1
		SOX4
		STAT1
		SYNGR1
		TCEAL1
		TFF3
		TP53BP2
		TYMS
		ZNF24

C

Perou et al., 2001			
Basal signature genes found in mRNA seq and		**ERBB2 signature genes found in mRNA seq and**	
TNBC vs Non-TNBC	**TNBC vs HER2-positive**	**Non-TNBC vs HER2-positive**	**TNBC vs HER2-positive**
Overlap	**Overlap**	**Overlap**	**Overlap**
ANXA1	ANXA1	ERBB2	ERBB2
DUSP6	CX3CL1	GRB7	GRB7
KRT17	KRT17	SMARCC2	SMARCC2
LAMC2	KRT5		
MMP14	LAMC2		

D

Sorlie et al., 2001		
TNBC vs HER2-positive genes found in mRNA seq and Microarray	**TNBC vs Non-TNBC genes found in mRNA seq and Microarray**	**Non-TNBC vs HER2-positive genes found in mRNA seq and Microarray**
CDH3	CDH3	ERBB2
CHI3L2	CHI3L2	GRB7
KRT17	CX3CL1	SMARCE1
	FABP7	TGFB1
	KRT17	
	KRT5	
	NFIB	
	SLPI	

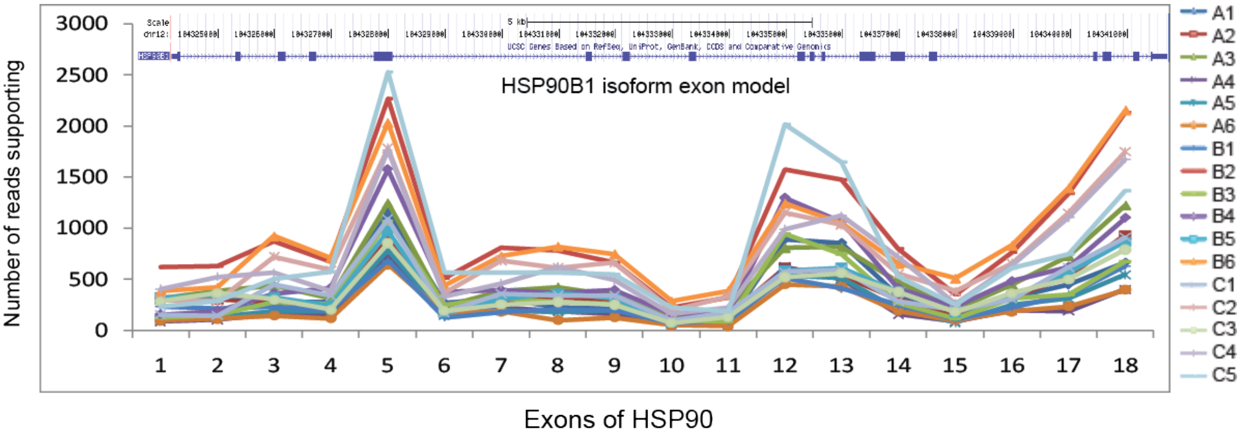
E

Sorlie et al., 2006	
TNBC vs HER2-positive genes found in mRNA seq and Microarray	**TNBC vs Non-TNBC genes found in mRNA seq and Microarray**
AGR2	ABAT
ANXA1	AGR2
CHST3	ANXA1
CREB3L4	CDK6
KRT23	CHST3
MLPH	CREB3L4
NMB	CTSL2
TFF3	DNAJC12
	EGFR
	EN1
	FABP7
	FBP1
	GABRP
	KRT23
	LMO4
	MLPH
	PSAT1
	S100A8
	S100B
	SCRG1
	SLC27A2
	SLPI
	SUSD3
	TFF3
	ZNF552

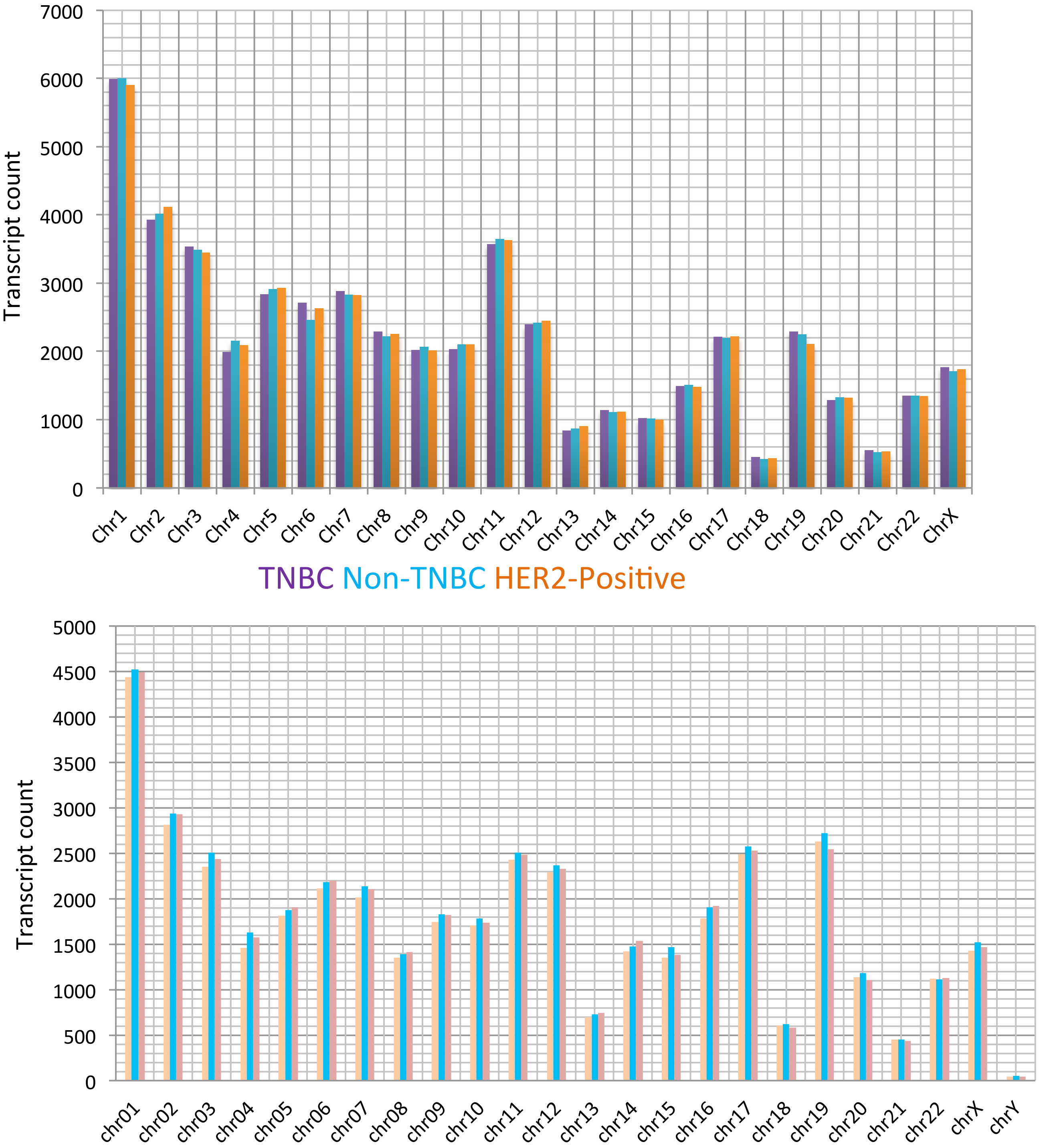
F

Parker et al., 2009		
Non-TNBC vs HER2-positive genes found in mRNA seq and Microarray	**TNBC vs HER2-positive genes found in mRNA seq and Microarray**	**TNBC vs Non-TNBC genes found in mRNA seq and Microarray**
BAG1	ACTR3B	ACTR3B
BCL2	BLVRA	ANLN
BLVRA	CCNE1	BAG1
CENPF	CDC20	BCL2
CXXC5	CDH3	BLVRA
ERBB2	ERBB2	CCNE1
ESR1	FOXA1	CDC20
GRB7	GPR160	CDH3
MDM2	GRB7	CENPF
MIA	KIF2C	CXXC5
MKI67	KRT17	EGFR
MLPH	MLPH	ESR1
MMP11	TMEM45B	FOXA1
NAT1		FOXC1
PGR		GPR160
SLC39A6		KRT14
		KRT17
		KRT5
		MDM2
		MIA
		MLPH
		NAT1
		PGR
		SFRP1
		SLC39A6
		TYMS

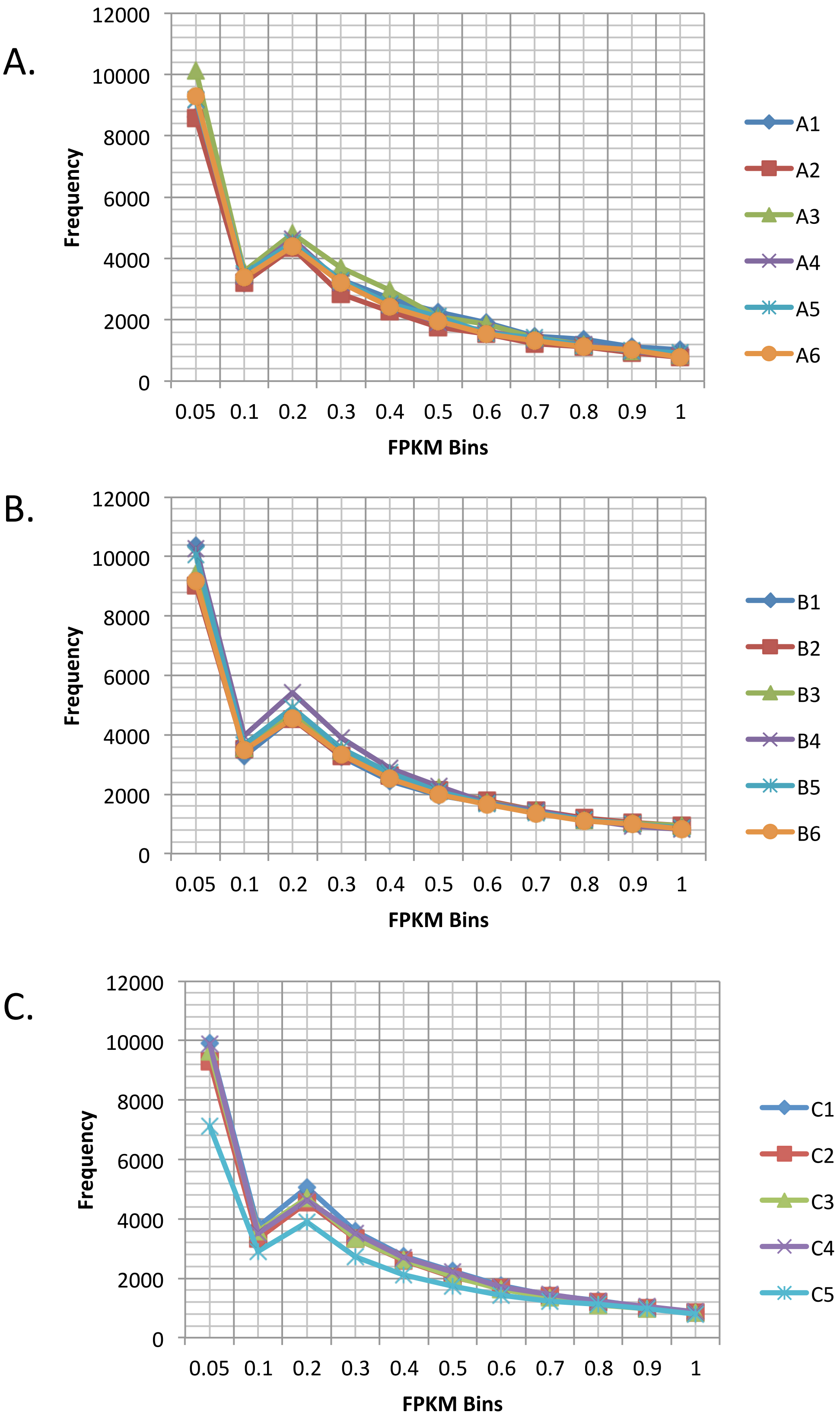
Supplement figure 1: The transcript model of HSP19B1 comprising 18 exons including the chromosome location from UCSC genome browser is shown in blue. The reassembled HSP19B1 complete transcript including all the 18 exons in all 17 samples shown in different colours.



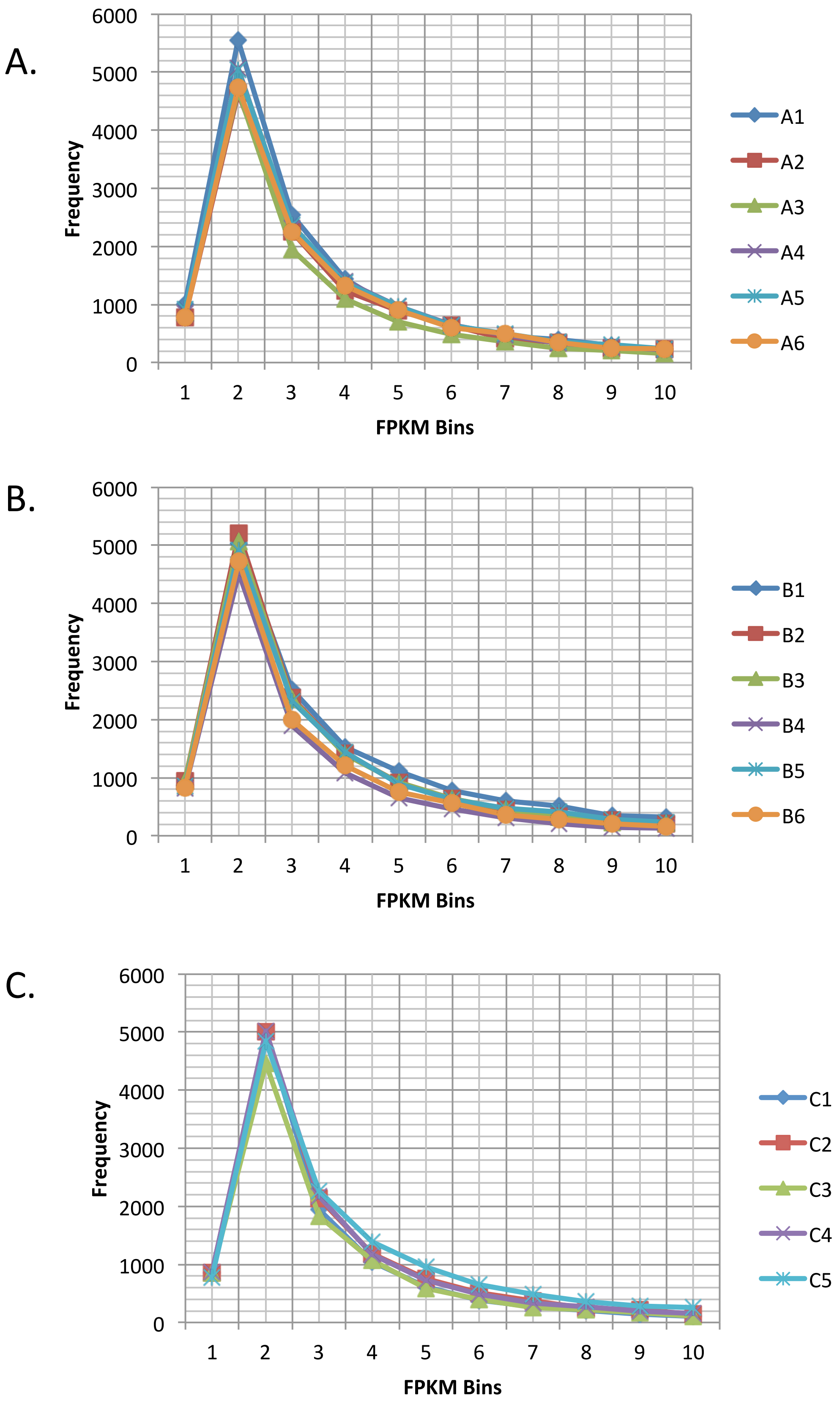
Supplement figure2: Transcripts distribution overview in each group. The average of expressed transcripts from each sample of the group was derived from cufflinks assembler using **A.** Ensembl GRCh37.62 **B.** hg19 UCSC database



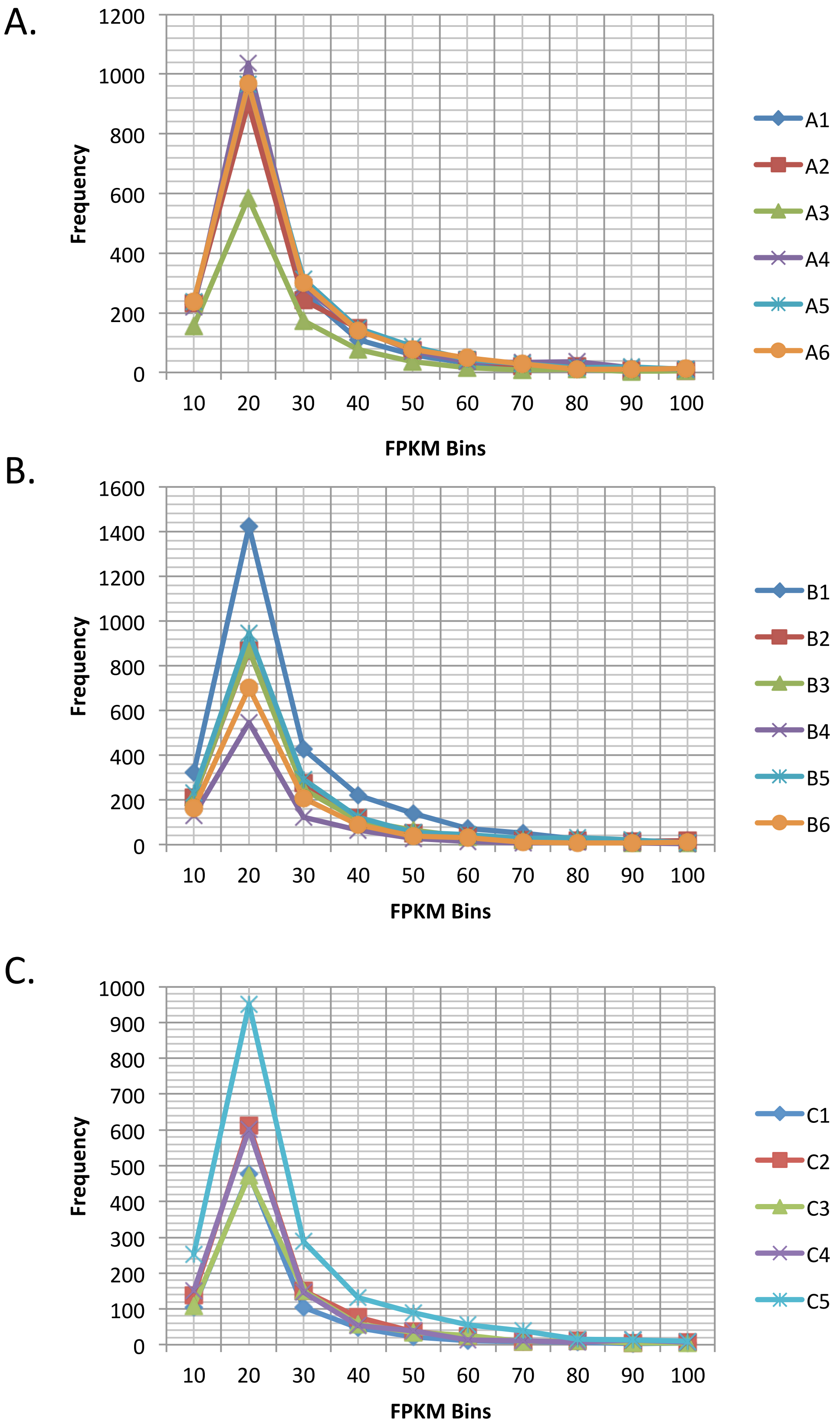
Supplement figure 3: FPKM Bins of low abundant transcripts reassembled from cufflinks assembler using Ensembl GRCh37.62 in **A.** TNBC **B.** Non-TNBC and **C.** Her2-Positive Expression Transcripts



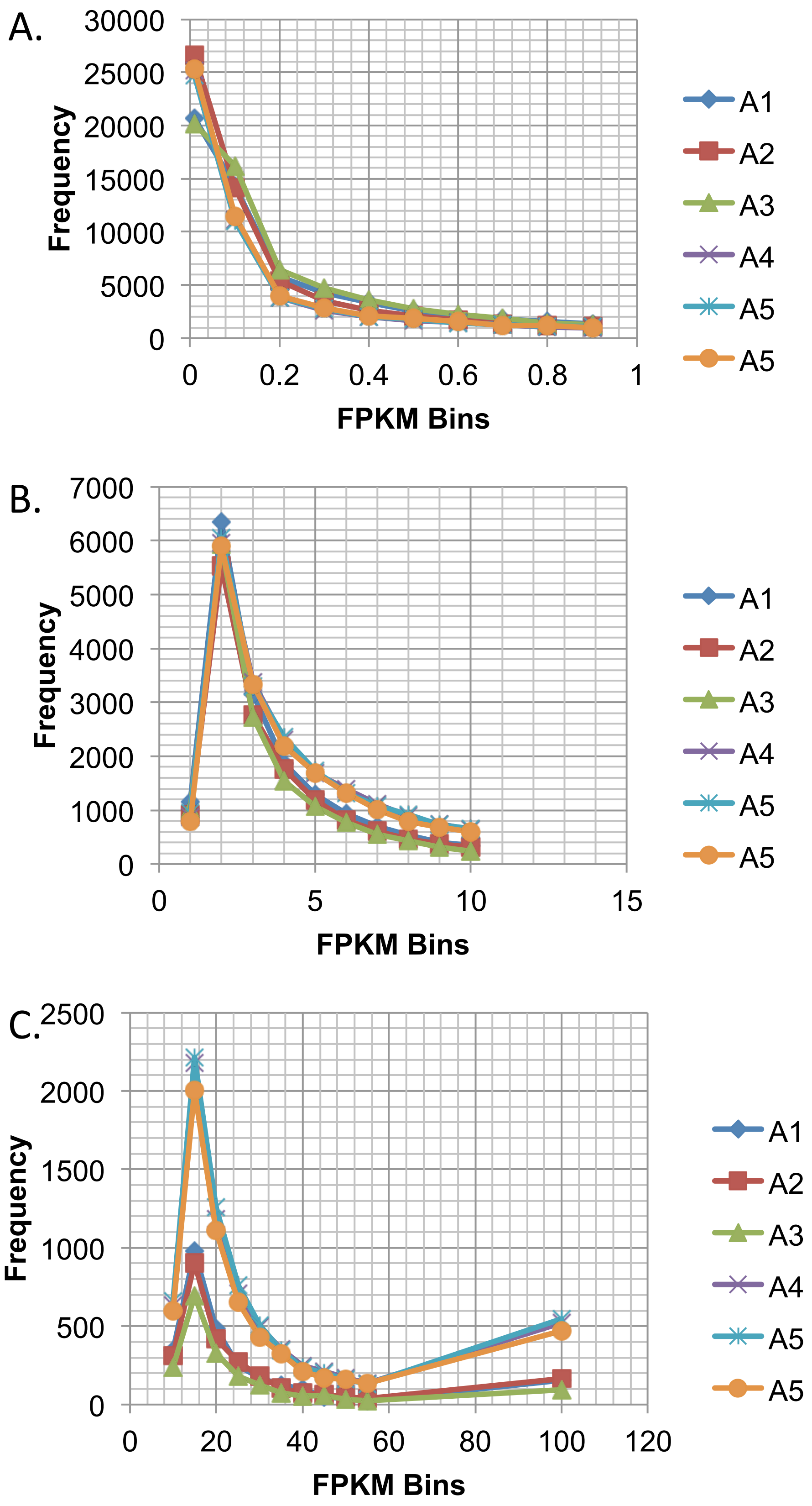
Supplement Figure 4: FPKM Bins of medium abundant transcripts reassembled from cufflinks assembler using Ensembl GRCh37.62 in **A.** TNBC **B.** Non-TNBC and **C.** Her2-Positive Expression Transcripts



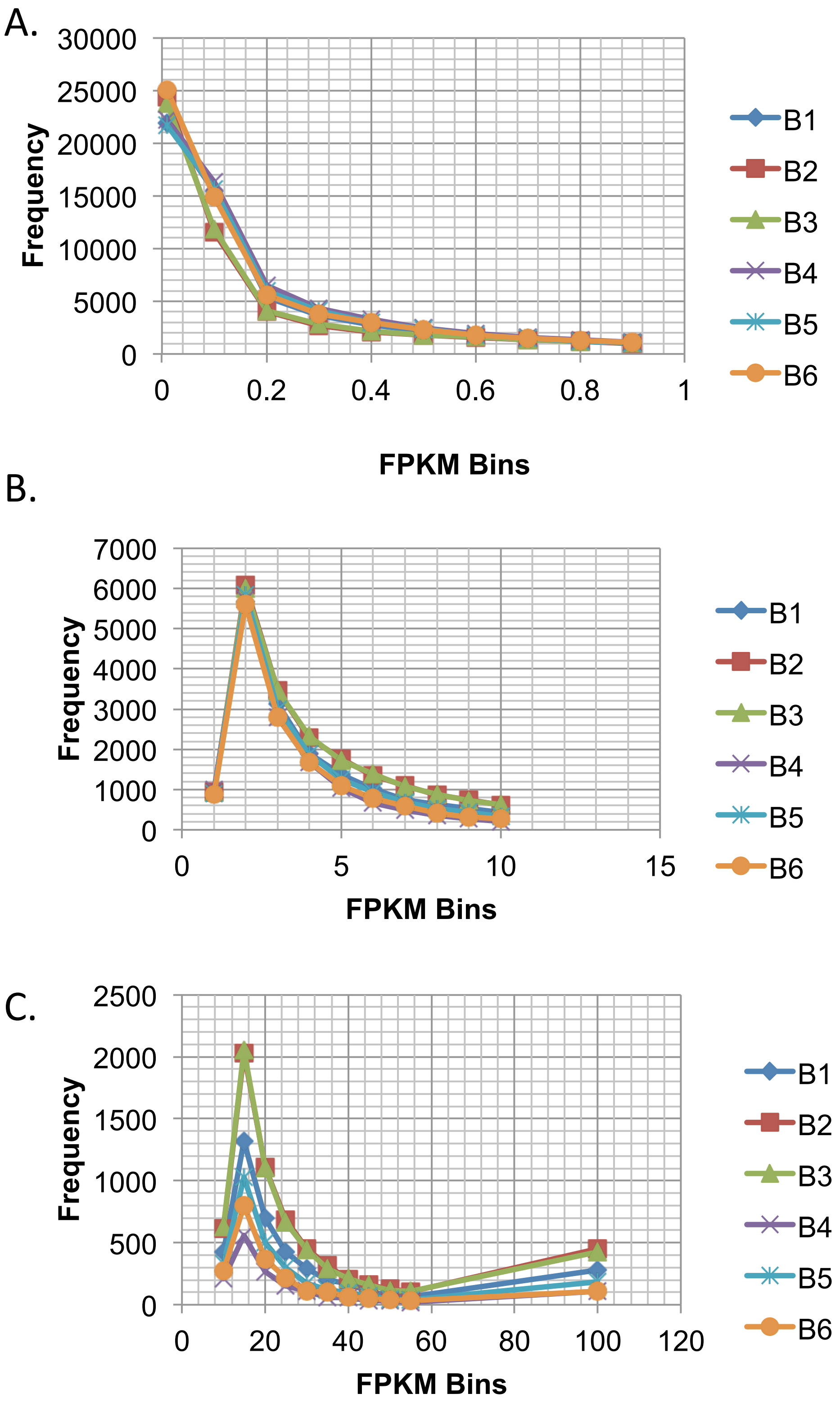
Supplement figure 5: FPKM Bins of high abundant transcripts reassembled from cufflinks assembler using Ensembl GRCh37.62 in **A.** TNBC **B.** Non-TNBC and **C.** Her2-Positive Expression Transcripts



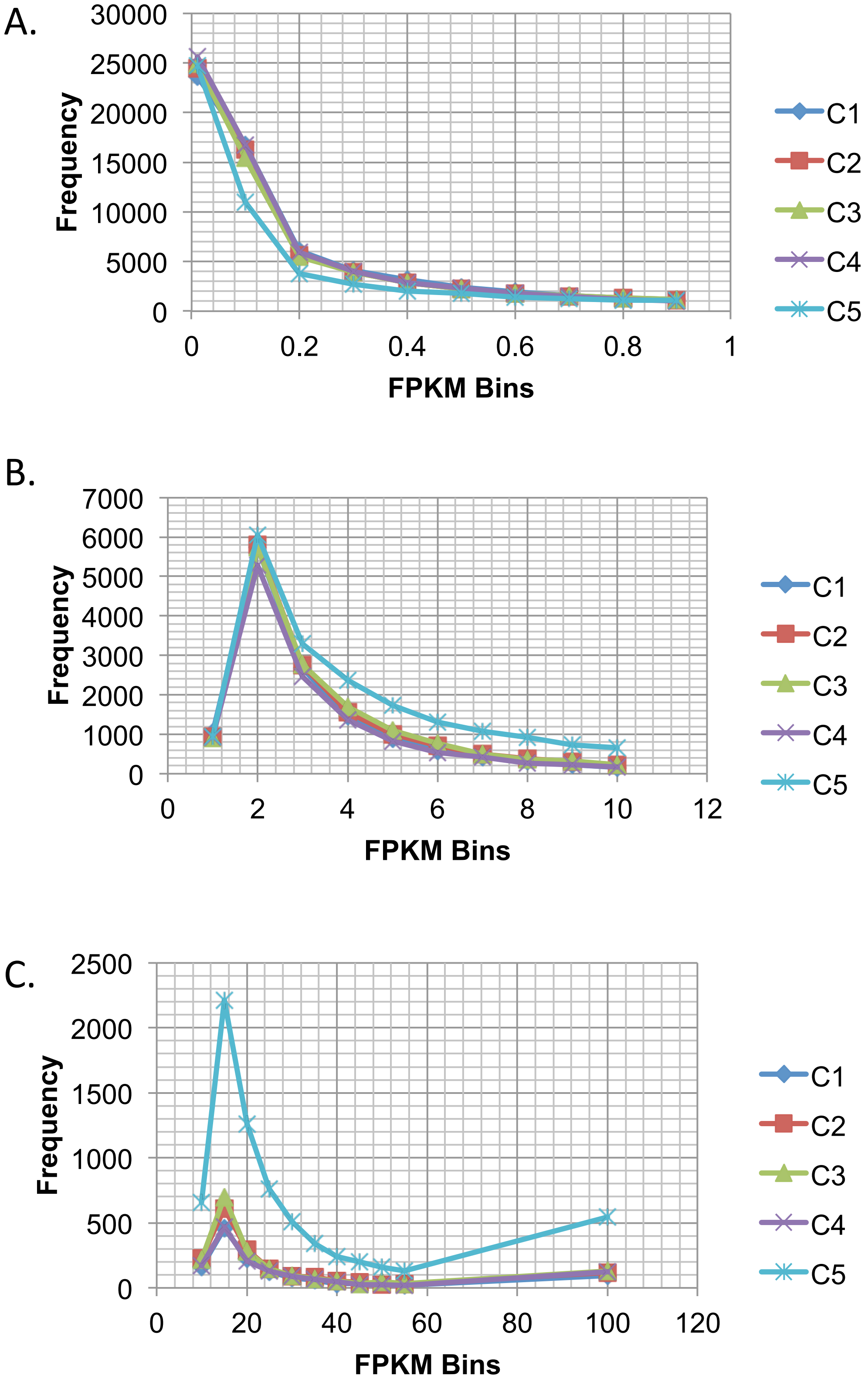
Supplement figure 6: FPKM Bins of low abundant transcripts reassembled from cufflinks assembler using UCSChg19 in **A.** TNBC **B.** Non-TNBC and **C.** Her2-Positive Expression Transcripts



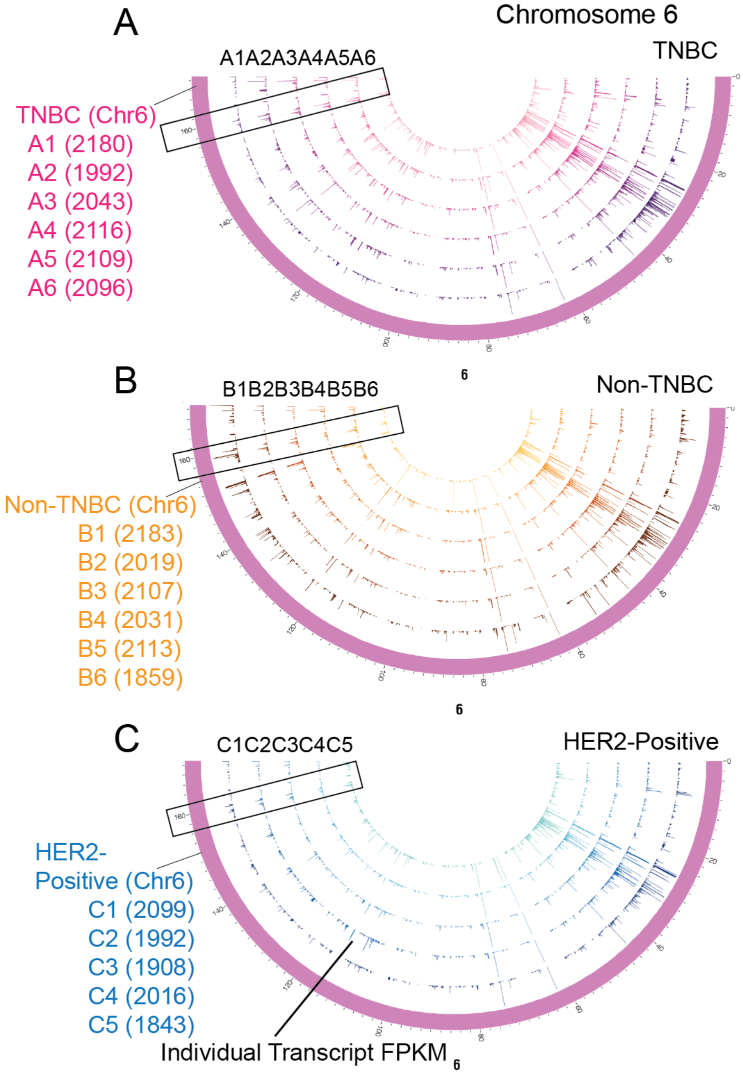
Supplement Figure 7: FPKM Bins of medium abundant transcripts reassembled from cufflinks assembler using UCSChg19 in **A.** TNBC **B.** Non-TNBC and **C.** Her2-Positive Expression Transcripts



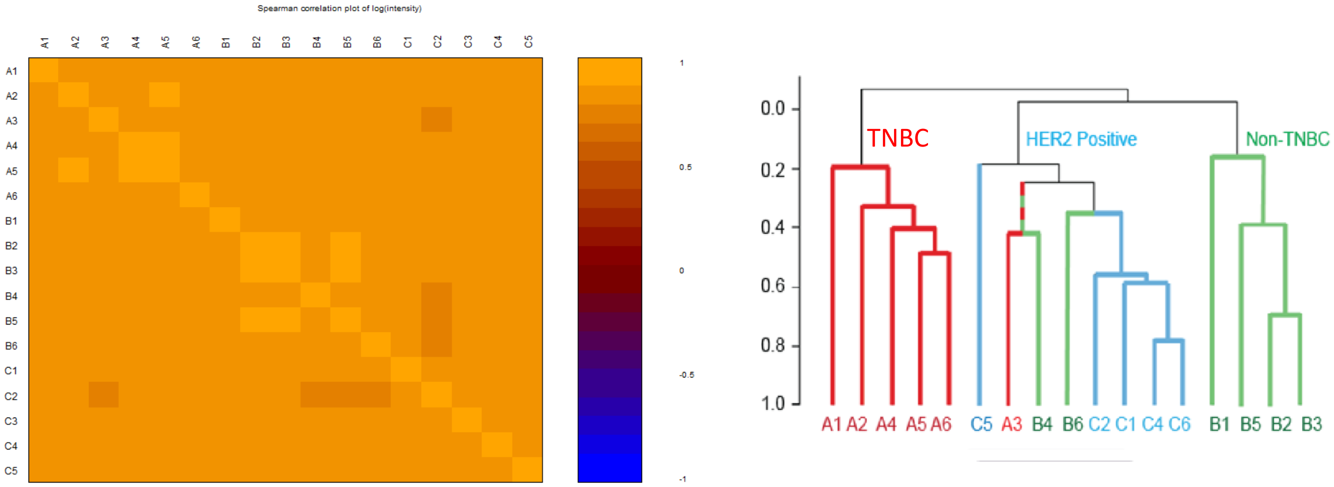
Supplement Figure 8: FPKM Bins of high abundant transcripts reassembled from cufflinks assembler using UCSChg19 in **A.** TNBC **B.** Non-TNBC and **C.** Her2-Positive Expression Transcripts



Supplement Figure 9: To highlight the scope of individual variations between the samples in each group the expression profile close up of chromosome 6 from all the six samples of **(A)** TNBC **(B)** Non-TNBC and **(C)** HER2 positive groups are shown as circos plot. The total number of transcripts (above FPKM 0.01) in chromosome 6 in each sample is provided in brackets next to the sample label.



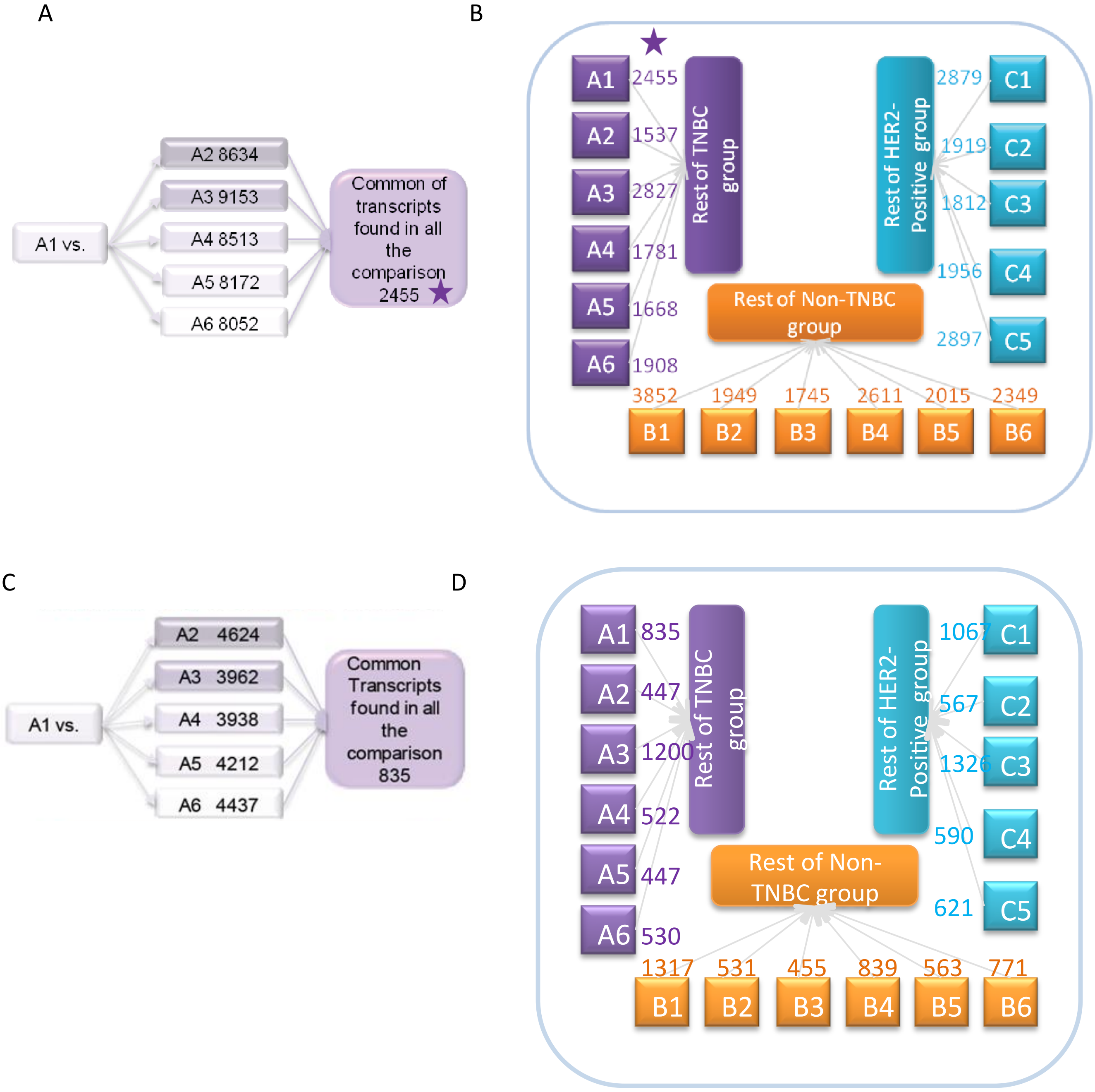
Supplement Figure 10: **(A)** Spearman correlation coefficient based heat map of all 17 samples based on gene expression. **(B)** Dendrogram showing the distance metric relationship between all of the samples. The correlations were computed separately for each pair of tissues and clustered according to similarity using average linkage hierarchical clustering.



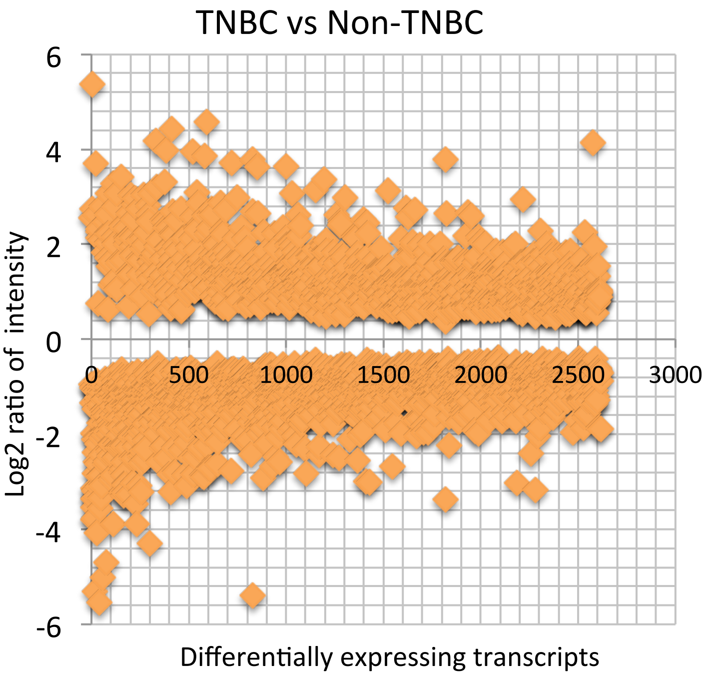
Supplement Figure 11: The schematic showing individual comparisons of the differentially expressing transcripts from

A. The cuffdiff statistical analysis of differentially expressed transcripts from the Ensembl (Figures AB) and UCSC reference (Figures CD) based transcripts assembly. Pairwise cuffdiff analysis of the transcript expression profile within each group identified high confidence statistically significant transcripts. As an example, the number of statistically significant transcripts expressed by the pairwise comparison of A1 versus the rest of the group is shown. From the pairwise comparisons of all of the samples within each group, the common transcripts present in all pairwise comparisons related to one specific sample against the rest of the group; i.e., 2455 transcripts in the case of A1 versus the rest of the group were detected.

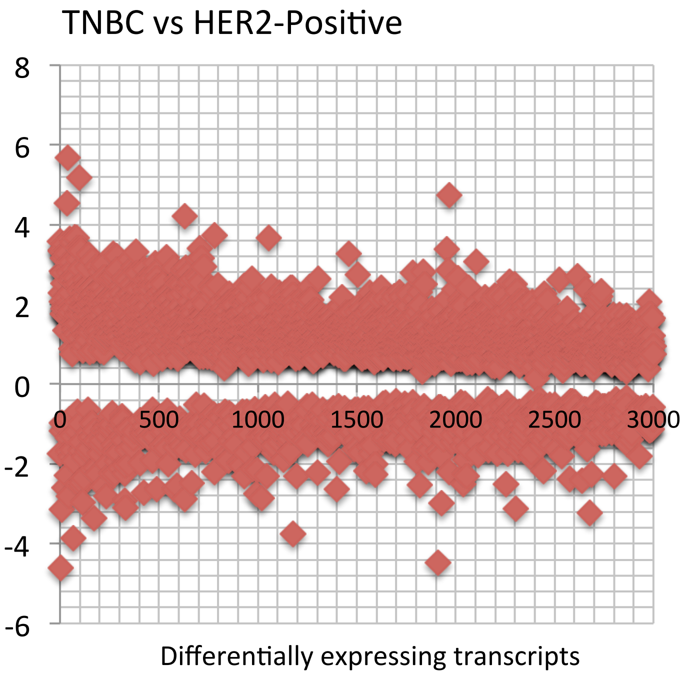
B. The above mentioned statistical analysis for the whole sample set is presented as a matrix. Higher number of common statistically significant transcripts indicates a possible divergence of that specific sample from the whole group, and these diverging outliers are coloured red.



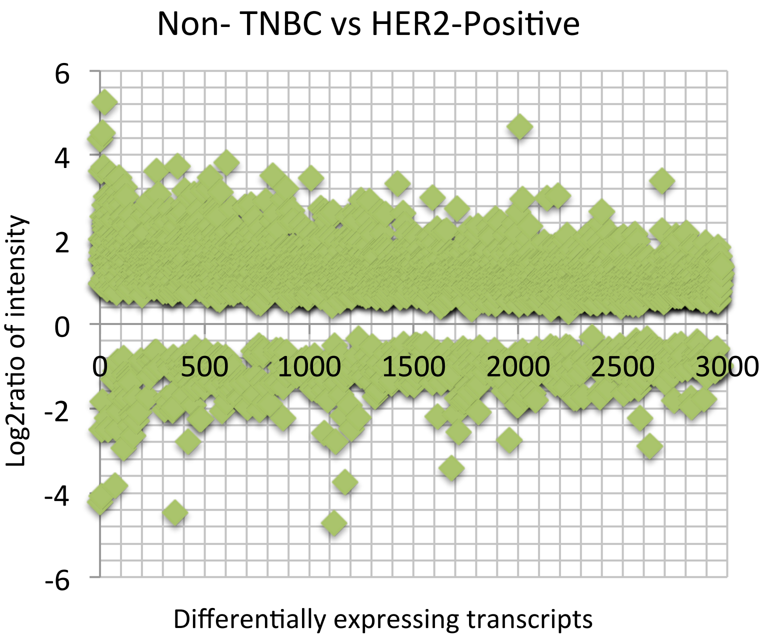
Supplement Figure 12 : Log2 ratio of intensity plot showing up and down regulated transcripts from differential expression analysis (supplement methods)



Up regulated in TNBC	962
Up regulatedin Non-TNBC	1655

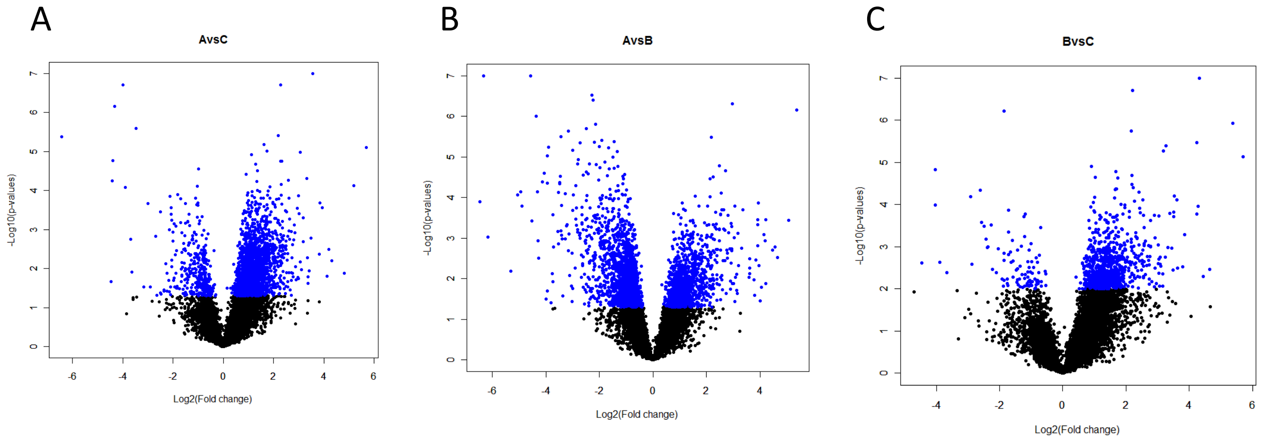


Up regulated in TNBC	2416
Up regulated in HER2-Positive	673

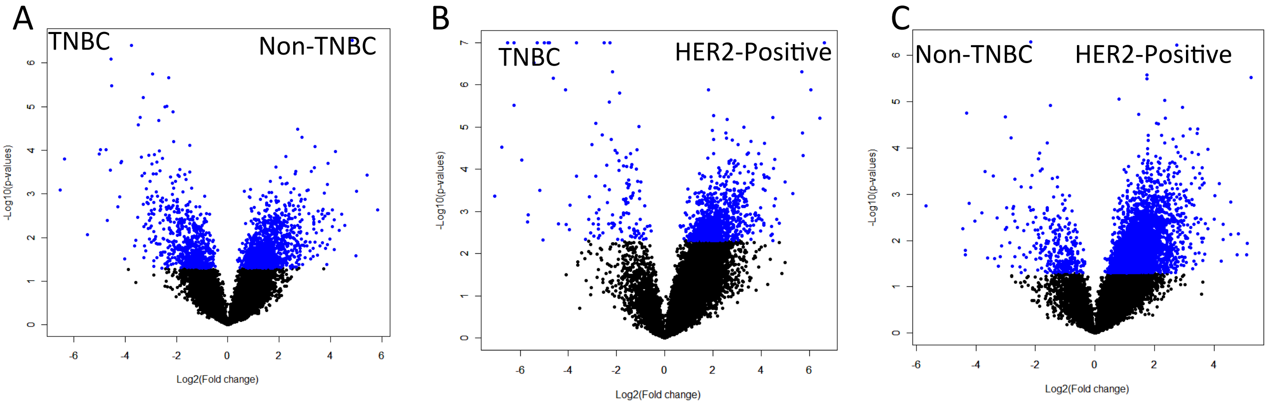


Up regulated in Non-TNBC	2509
Up regulated in HER2-Positive	463

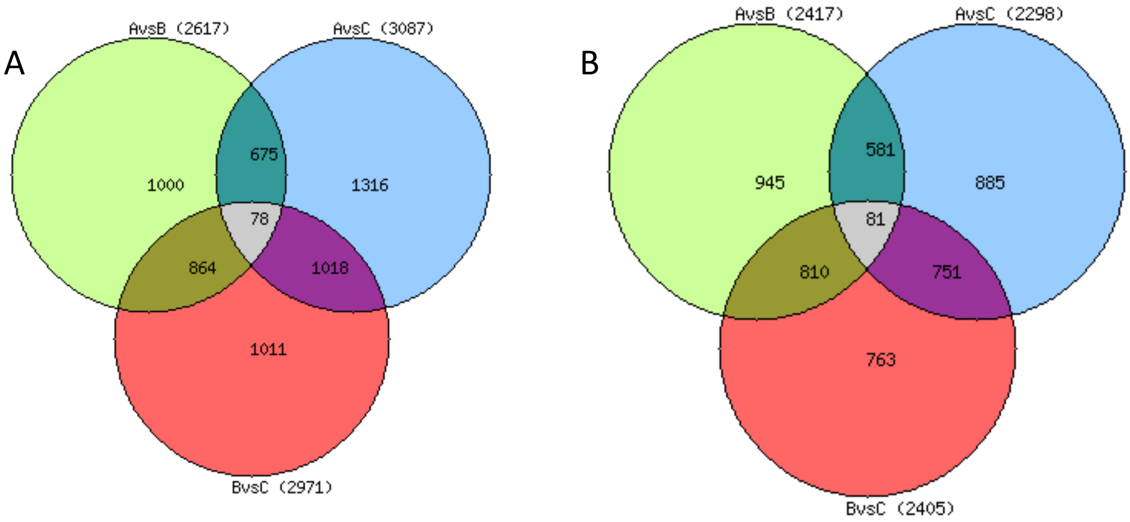
Supplement Figure 13: Volcano plot showing differentially expressing statistically significant genes (A, B and C) assembled based on genome Ensembl GRCh37.62 from F-test of TNBC vs Non-TNBC, Non-TNBC vs HER2 Positive and TNBC vs HER2-Positive genes and transcripts



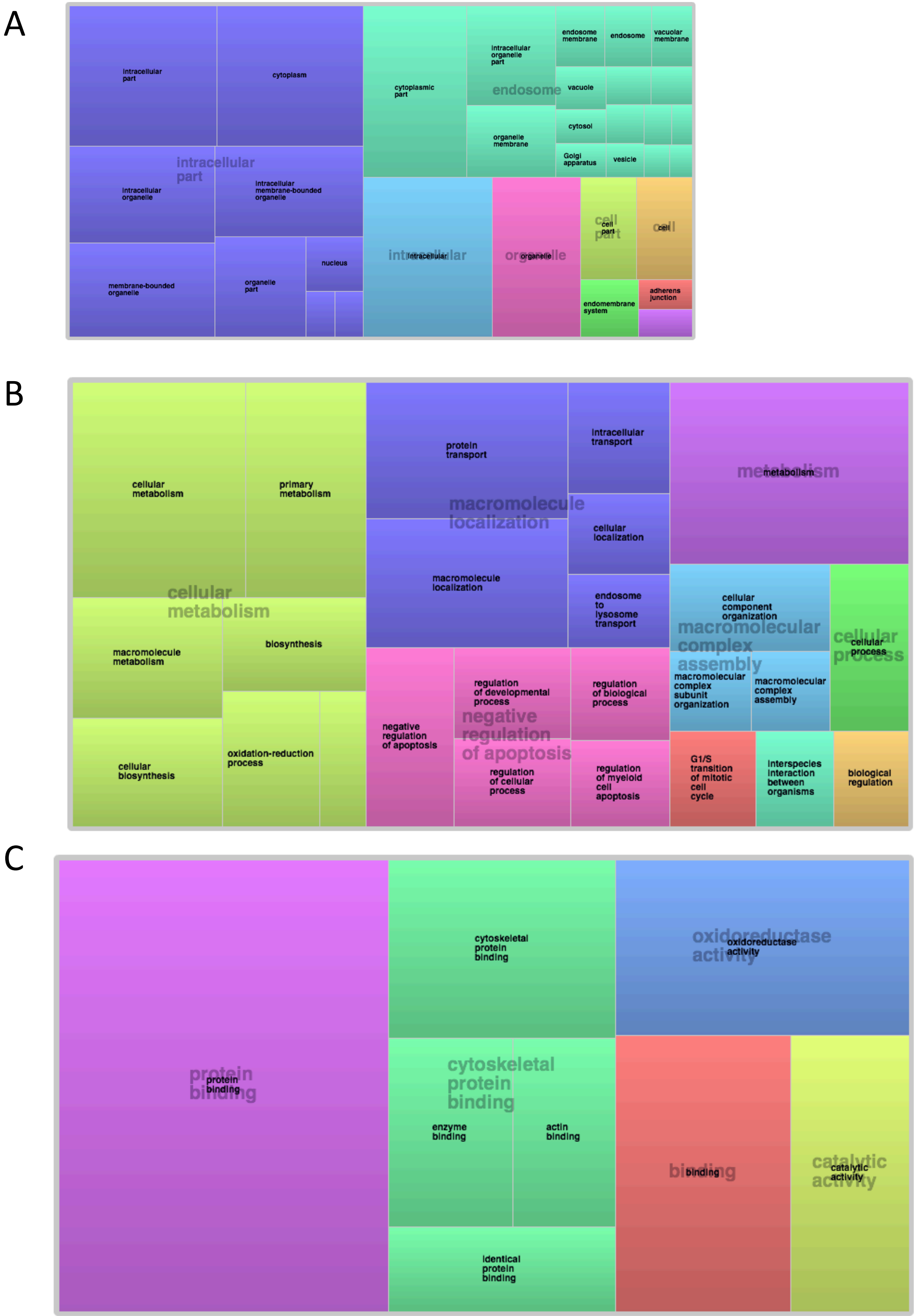
Supplement Figure 14: Volcano plot showing differentially expressing statistically significant transcripts (A, B and C) (assembled based on UCSC HG19 genome) from F-test of TNBC vs Non-TNBC, Non-TNBC vs HER2 Positive and TNBC vs HER2-Positive genes and transcripts



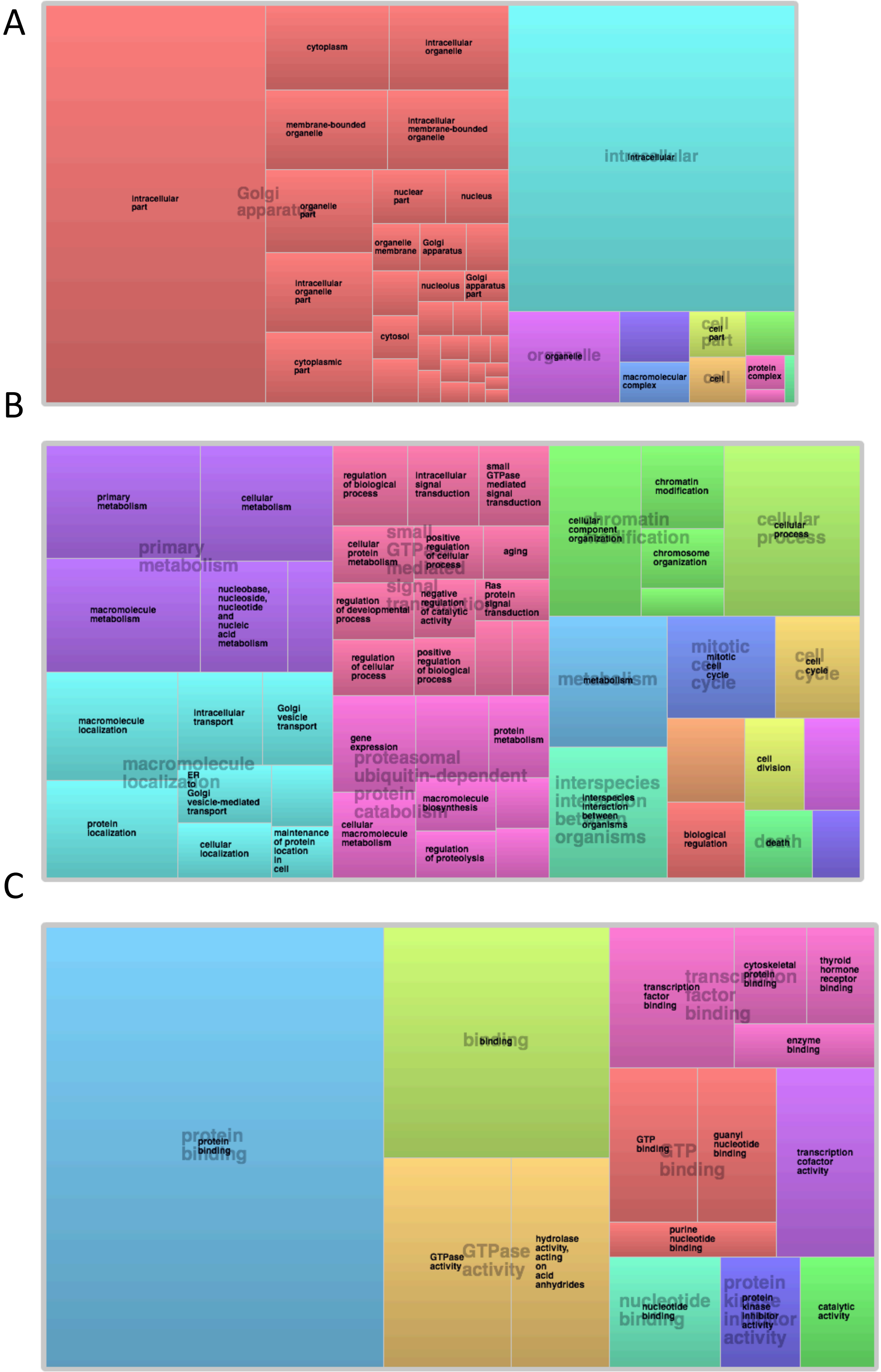
Supplement Figure 15: Venn Diagram showing differentially expressing statistically significant transcripts (A) genes (B) assembled based on genome Ensembl GRCh37.62 from F-test of TNBCvs Non-TNBC (AvsB), Non-TNBC vs HER2 Positive (BvsC) and TNBC vs HER2-Positive genes and transcripts



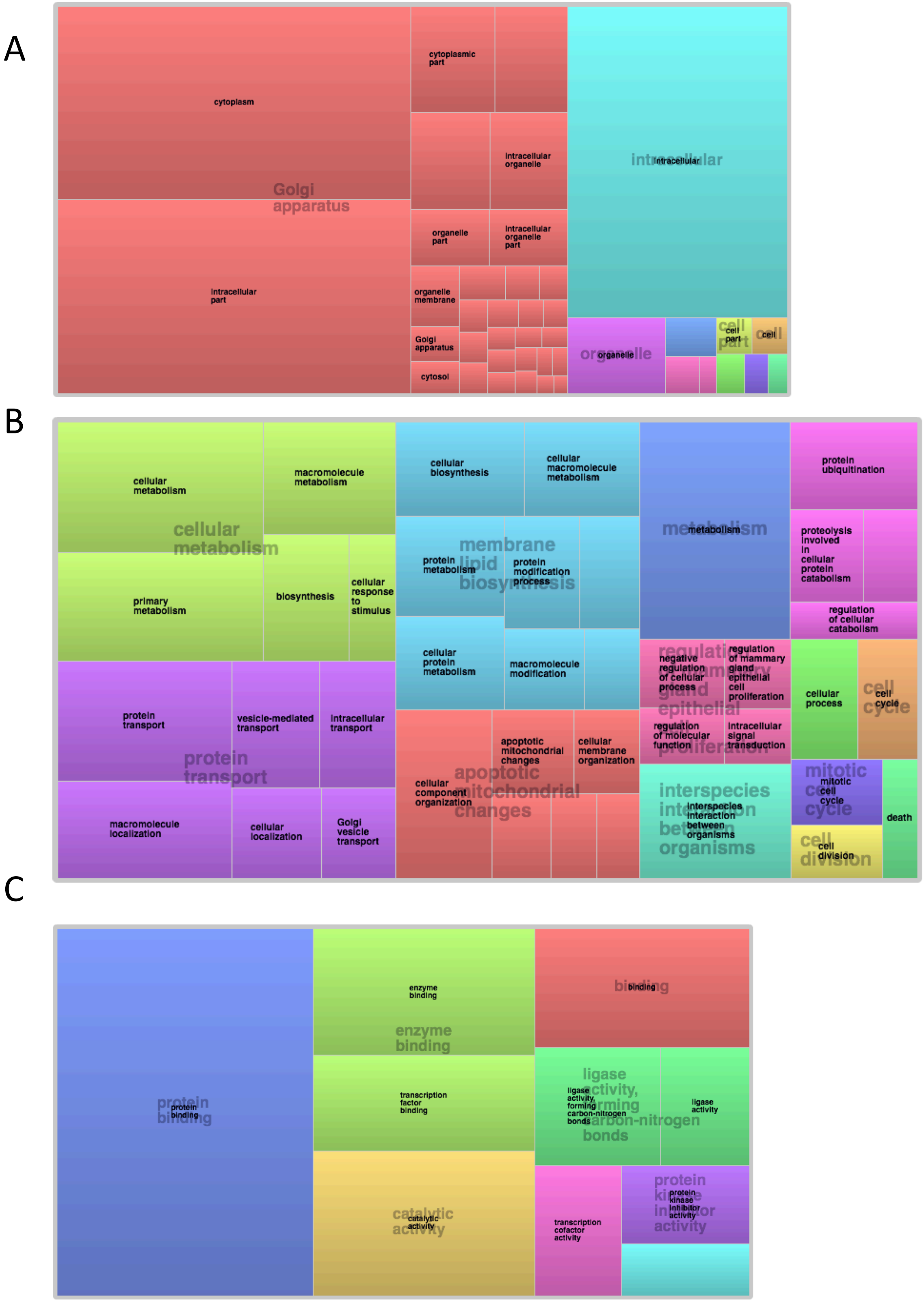
Supplement Figure 16 : GOAnalysis of TNBC vs Non-TNBC differentially expressing transcripts assembled using Ensembl GRCh37.62 as a reference (A)cellular, (B) biological and (C) molecular functions



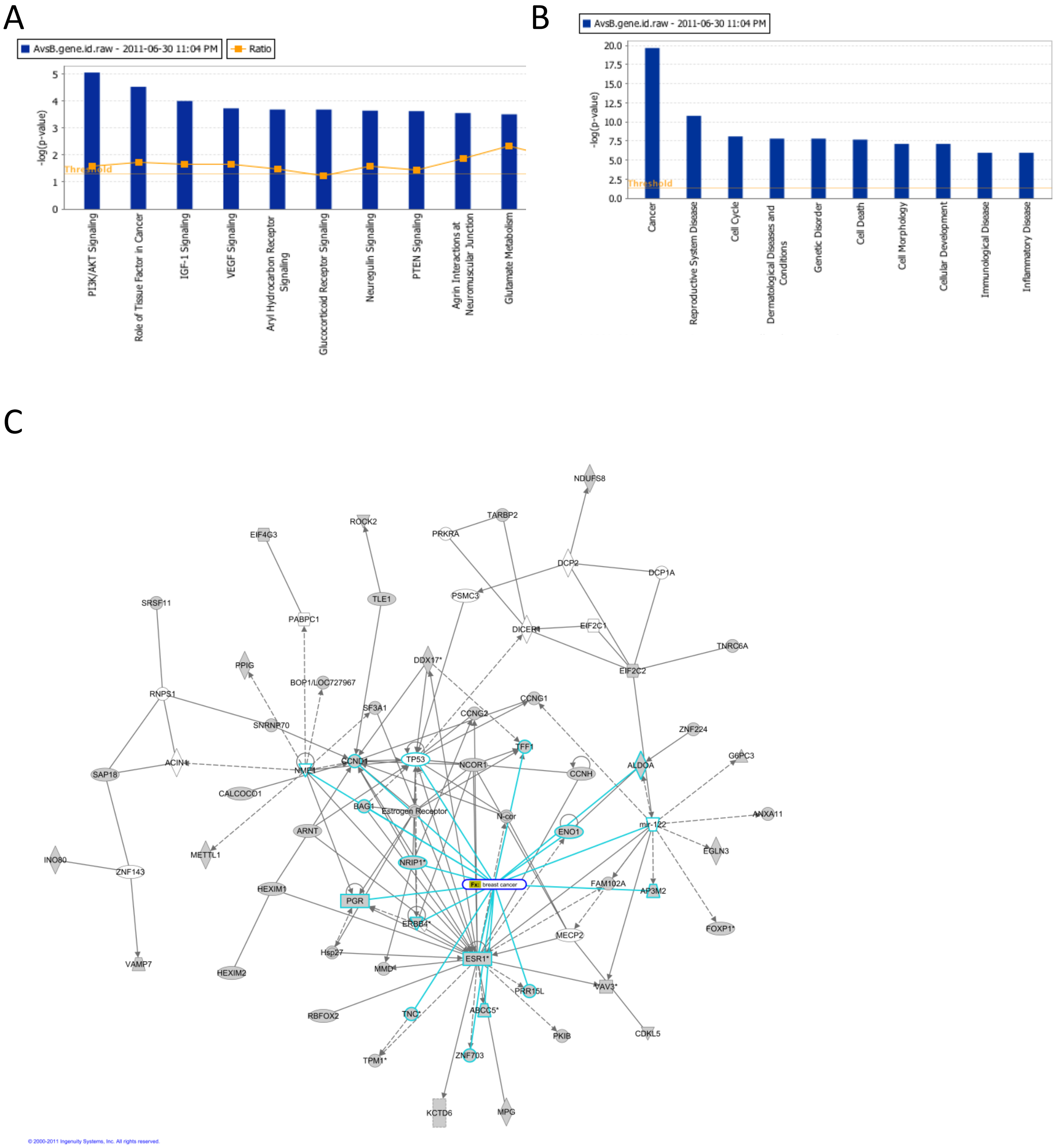
Supplement Figure 17 : GO Analysis of TNBC vs HER2-Positive differentially expressing transcripts assembled using Ensembl GRCh37.62 as a reference (A)cellular, (B) biological and (C) molecular functions



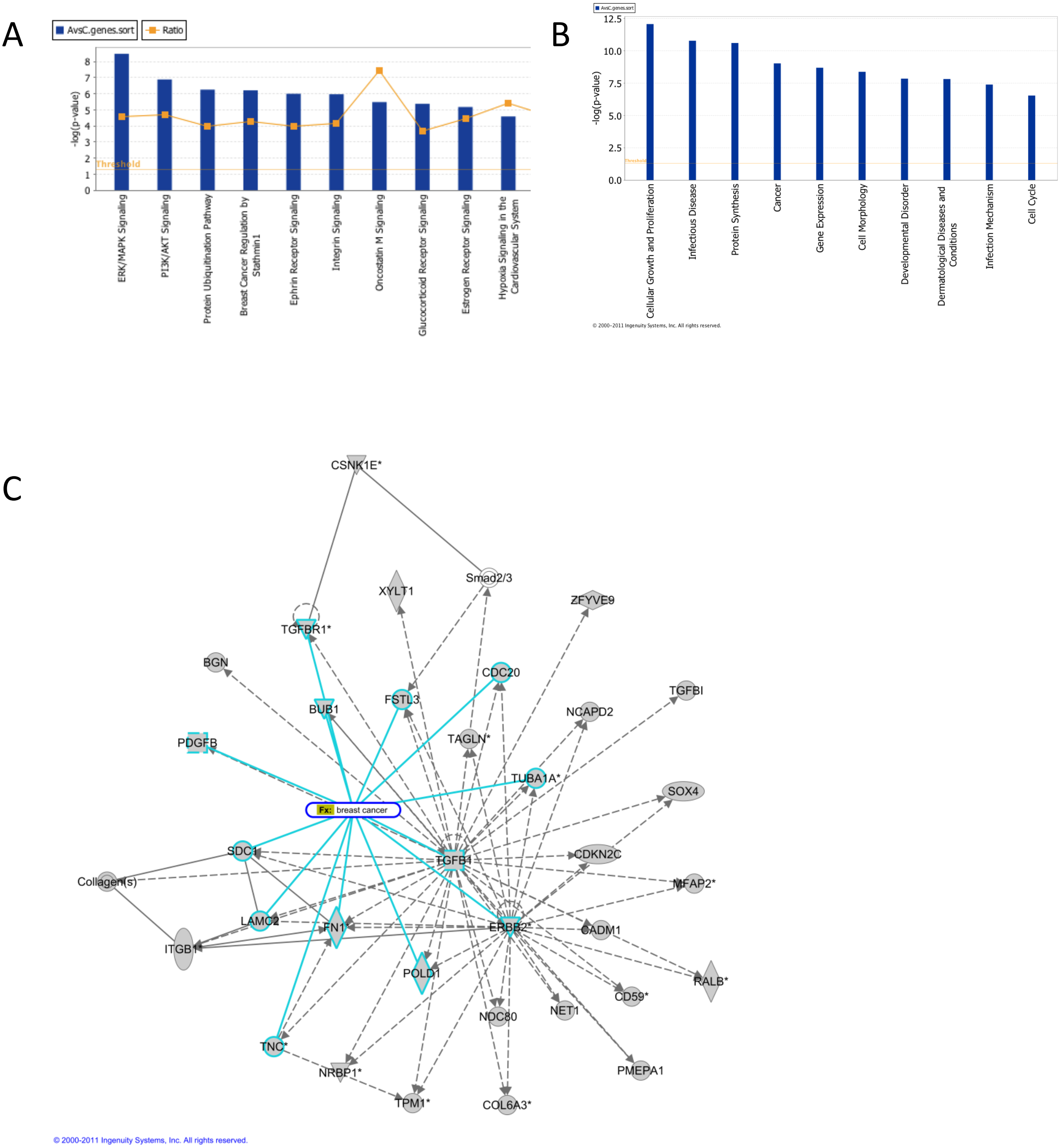
Supplement Figure 18: GO Analysis of Non-TNBC vs HER2-Positive differentially expressing transcripts assembled using Ensembl GRCh37.62 as a reference (A)cellular, (B) biological and (C) molecular functions



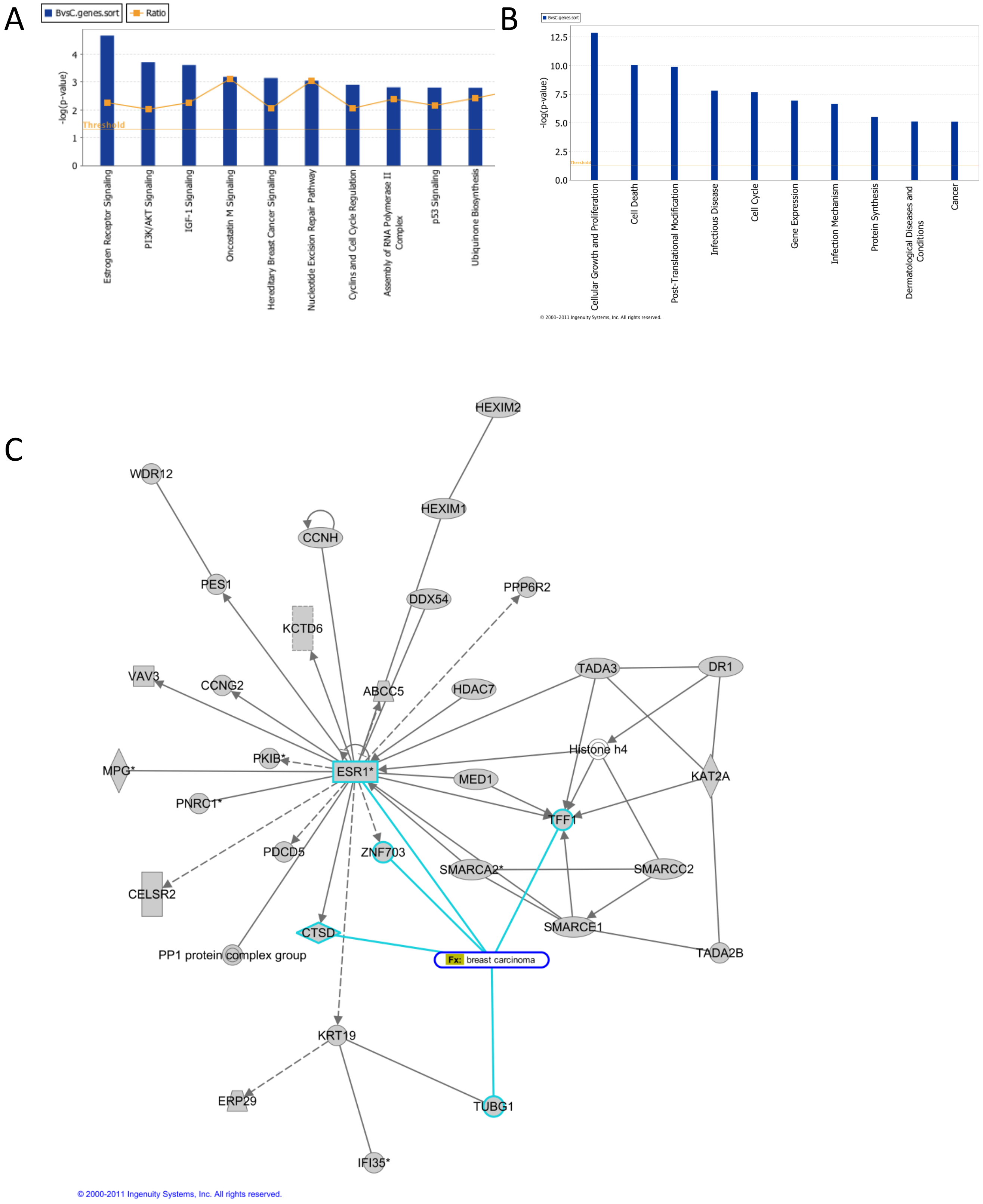
Supplement Figure 19 : IPA Analysis showing (A) canonical (B) functional and (C) signaling network influenced by TNBC vs Non-TNBC differentially expressing Transcripts



Supplement Figure 20 : IPA Analysis showing (A) canonical (B) functional and (C) signaling network influenced by TNBC vs HER2-Positive differentially expressing Transcripts assembled using Ensembl



Supplement Figure 21: IPA Analysis showing (A) canonical (B) functional and (C) signaling network influenced by Non-TNBC vs HER2-Positive differentially expressing transcripts



Supplement Figure 22: The schematic showing the strategy of microarray analysis and how the genes identified from analyses microarray are compared against the mRNA sequencing pair-wise comparisons