**Supplemental Files**

**Supplemental File 1. Differentially Expressed Genes between ICR4 and ICR1 samples.** A) Differentially expressed genes between TCGA ICR1 and ICR4 (q value < 0.05 and absolute log FC > 0.5). B) ICR cluster assignments.

**Supplemental File 2. Bindea Cell Signatures.** The difference of the enrichment score between ICR1 and ICR4 is shown with the corresponding FDR-adjusted *p* values (q values).

**Supplemental File 3. Differentially Mutated Genes between ICR1 and ICR4.**

**Supplemental File 4. GISTIC Regions.** (A) and (B) display the regions amplified and deleted respectively, with their peak identifier, the location by genome version hg19 the peak location by base wide and regular, the location of the region by base and the q-value. (C) displays the specific genes for the locations and which genes are immune related according to Immunome DB. (D) focuses on those regions that are specifically altered between ICR1 and ICR4.

**Supplemental File 5. Differentially Altered Genes between ICR1 and ICR4.** (A) and (B): genes with significantly higher gains or losses in ICR4 over ICR1, respectively. (C) and (D): genes with significantly higher gains or losses in ICR4 over ICR1, respectively. (E), (F), and (G): same data are shown for those genes that are immune related according to Immunome DB. No ICR1 gains containing immune-related genes were found. (H): summary of these findings.

**Supplemental Figures Legends**

**Supplementary Figure 1. Calinski & Harabasz Plot of the TCGA breast cancer dataset.** Plot generated with custom R functions.

**Supplementary Figure 2. IPA of the top deferentially expressed pathways between ICR 1 and 4.** Fold change between ICR1 and ICR4 was calculated using the edgeR R package. Results were uploaded into IPA and analysis was performed using a cut off of absolute log FC > 1 and a FDR-adjusted *p* value (q value) < 0.05. This resulted in 1100 upregulated genes and 998 downregulated genes.

**Supplementary Figure 3. Association between ICR genes, ICR Z-score, ICR cluster and overall survival. (A)** Hazard ratios were calculated comparing the lowest expressing quartile, tertile or cluster compared to the highest (TCGA RNASeq samples). Hazard-ratios (HR) were determined by Cox proportional hazards regression model (95% confidence interval); *p* values are from log-rank test. (B) Expression values of the 20 signature genes across the four ICR subtypes, plus CXCL11 and the ICR-score (average of the 20 signature genes per sample).

**Supplementary Figure 4. Consensus clustering of the breast-cancer validation dataset.** (A) Consensus cluster matrix generated by consensus cluster plus R package, repeats = 5000, Kmax =7 and agglomerative hierarchical clustering with ward criterion (Ward.D2) inner and complete outer linkage. (B) Micro-array expression heatmap using gplots N = 1954; (19 Genes, 28 Probes). (C) Calinski & Harabasz test validation micro-array dataset.

**Supplementary Figure 5. Mutational load, neoepitope load, and immune phenotypes.** (A) Mutation frequencies (B) Number of strong epitopes (C) Number of mutation yielding strong neoepitopes, by ICR cluster: IMS subtype and ICR cluster within different subtypes. N = 904, samples with mutation, neoepitope or neoepitope yielding mutation number of > 250, > 40, and > 25 respectively have not been plotted but are included in the statistics, *p* values determined by Kruskal-Wallis test. Labels in the y axis represent median and average number per sample on the top and bottom respectively.

**Supplementary Figure 6. TP53 mutations, MAPK mutations, ICR and survival.** A) Kaplan-Meier overall survival curves of TCGA breast cancer RNA-seq dataset based on TP53 and MAPK (MAP2K4 or MAP3K1) mutational status. B) Kaplan-Meier overall survival curves of TCGA breast cancer RNA-seq dataset based on TP53 and MAPK (MAP2K4 or MAP3K1) mutational status in ICR1 and 4 subtypes. Hazard-ratios (HR) were determined by Cox proportional hazards regression model (95% confidence interval, panel A); *p* values are from log-rank test.

**Supplementary Figure 7. Landscape of copy number alteration by IMS within each ICR group.**  Segmented copy number data of 995 TCGA primary breast cancer samples from TCGA grouped by ICR immune subtype (first annotation row) and IMS clinical type (second sample annotation row).  Red shading indicates the degree of copy gain/amplification of a region; blue shading indicates copy loss/deletion; IMS:  intrinsic molecular subtype.

**Supplementary Figure 8. *CXCL9*, *CXCL10*, and *CXCL11* expression, amplification, and association with overall survival.**

A) Proportion of TCGA breast cancer samples bearing *CXCL9*, *CXCL10*, and *CXCL11* amplifications and deletions in each ICR subtype.; *q* value is from Chi-square test, ICR1 amp vs ICR4 amp. B) Expression of *CXCL9*, *CXCL10*, and *CXCL11* according to the copy number variation status in TCGA breast cancer samples; *p* values are from t-test. C) Kaplan-Meier overall survival curves of TCGA breast cancer RNA-seq dataset based on *CXCL9-11* copy number status in all samples (left panel) and ICR1 and ICR4 samples (right panel). Hazard-ratios (HR) were determined by Cox proportional hazards regression model (95% confidence interval, panel A); *p* values are from log-rank test.

*CXCL9, CXCL10*, and *CXCL11* were always co-amplified or co-deleted and therefore are all represented vu the same graph in panel A and C.

**Supplementary Figure 9. The MAPK-mutation score can segregate different immune phenotypes of breast cancer across intrinsic molecular subtypes.** Left panel: MAPK-pathway genes differentially expressed between *MAP3K1* or *MAP2K4* mutated (MAPK-mut) and wild-type TCGA Luminal samples are used to segregate ICR1-4 TCGA samples (Luminal, HER2-enriched and Basal-like subtypes, N=309); Right panel: MAPK-mut transcripts defined in the TCGA dataset are used to segregate ICR1-4 samples of the validation dataset (Luminal, HER2-enriched and Basal-lije subtypes, N=677). Samples are ordered by MAPK-mut score which is the average ranking of the samples in up and down regulated z-scores (see methods for detail).

The TCGA heatmaps are based on the TCGA samples for which mutational data were available.

**Supplementary Figure 10. Expression boxplots of the up and down regulated genes of the MAPK pathway according to ICR status.** Genes identified as part of the MAPK pathways and as differentially expressed between *MAP3K1* or *MAP2K4* mutated and wild-type samples; RNAseq data has been normalized and log2 transformed (see method section).