Supplemental Data - Tables

Chemical composition is maintained in poorly conserved intrinsically disordered regions and suggests a means for their classification - Moesa et al.

Tables

Table S1. Average values of residue, residue type and type content conservation used as cutoffs for the definition of IDR groups.

Average Value	**Predicted (Disodb)**	**Predicted (IUPred)**	**Experimental (DisProt)**
Residue conservation score	6.537	6.426	7.465
Residue type conservation score	7.123	6.976	7.919
Type content conservation score*	0.105	0.115	0.059
HR Type conservation score + 2 standard deviations †	0.143	0.155	0.077

* Type content conservation score is the Euclidean distance between the type content of an IDR in a human protein with that in a corresponding aligned sequence in orthologous proteins. A smaller score indicates greater conservation of chemical composition.

† Average type content conservation score + 2 standard deviations of IDRs with high residue (HR) conservation. This value was used to define IDRs with low residue and high type content conservation.

Table S2. Number of IDRs within each sub group as determined by the levels of residue, type and type content conservation in all IDR datasets.

Intrinsically disordered regions	**Predicted (Disodb)**	**Predicted (IUPred)**	**Experimental (DisProt)**
Total	14612	9464	102
High residue conservation (HR)	8075	5238	58
Low residue conservation (LR)	6537	4226	44
Low residue and low type conservation (LRLT)	5764	3739	39
Low residue, low type and high type content conservation (HTC)	3042 (53%)	1910 (51%)	20 (51%)
Low residue, low type and low type content conservation (LTC)	2722	1829	19

Table S3. Representative GO terms enriched in IDRs classified into clusters based on chemical composition ($p<0.01$).

Term	Positive	Negative	Polar	Hydrophobic	Special
GOBP	RNA processing DNA packaging	Protein folding	Cell surface receptor linked signal transduction	Actin filament-based process	Ectoderm development
GOMF	RNA binding Helicase activity	Unfolded protein binding	Cytokine binding Peptide receptor activity	Lipid binding	Extracellular matrix structural constituent
GOCC	Protein-DNA complex Ribonucleoprotein complex		Intermediate filament	Ion channel complex	Collagen Proteinaceous extracellular matrix

Table S4. Top 5 Gene Ontology Biological Process terms enriched in IDRs separated by chemical composition and location ($p < 0.01$).

IDR cluster	**GO BP terms associated with IDR location**		
	N-terminal	**Middle**	**C-terminal**
Positive	NLS-bearing substrate import into nucleus	Regulation of dopamine secretion	RNA processing
	Base-excision repair	Chromatin modification	mRNA processing
	DNA packaging	Chromatin organization	mRNA metabolic process
	Protein import into nucleus	Chromosome organization	RNA splicing
	Chromatin assembly	Cell-cell signaling	Small GTPase mediated signal transduction
Negative	Amino acid transport	Carbohydrate transport	Protein folding
	Carboxylic acid transport	Transmembrane transport	Regulation of vesicle-mediated transport
	Organic acid transport	Negative regulation of protein kinase activity	Regulation of cell cycle
	Amine transport	Negative regulation of kinase activity	Regulation of apoptosis
		Negative regulation of transferase activity	Response to drug
Polar	Ubiquitin-dependent protein catabolic process	Pattern specification process	G-protein coupled receptor protein signaling pathway
		Embryonic pattern specification	Elevation of cytosolic calcium ion concentration
		Regionalization	Cell surface receptor linked signal transduction
		Protein amino acid phosphorylation	Cytosolic calcium ion homeostasis
		Regulation of binding	Cellular calcium ion homeostasis

Hydrophobic	Neurotransmitter transport	Vesicle-mediated transport	Intracellular signaling cascade
	Neutral amino acid transport	Protein amino acid phosphorylation	Small GTPase mediated signal transduction
	Branched chain family amino acid metabolic process	Membrane organization	Cell surface receptor linked signal transduction
	Phosphorylation	Immune effector process	Protein transport
	Glycine transport	Positive regulation of neurogenesis	Ion transport
Special	Amino acid transport	Cell adhesion	Response to abiotic stimulus
	Serine family amino acid catabolic process	Biological adhesion	Organ regeneration
	Posttranscriptional regulation of gene expression	Negative regulation of myeloid leukocyte differentiation	
	Amine transport	Regulation of myeloid leukocyte differentiation	
		Regulation of myeloid cell differentiation	

Table S5. GO terms enriched in HR clusters by type content ($p < 0.01$)

Term	**Positive**	**Negative**	**Polar**	**Hydrophobic**	**Special**
GOBP	-RNA processing -ribonucleoprotein complex biogenesis -histone acetylation -spliceosome assembly	-protein modification by small protein conjugation or removal	-morphogenesis of an epithelium -neural tube formation	-cognition -exocytosis	-ectoderm development -epidermis development
GOMF	-histone methyltransferase activity	-ubiquitin-protein ligase activity		-small GTPase regulator activity -lipid binding	-extracellular matrix structural constituent -ligand-dependent nuclear receptor transcription coactivator activity
GOCC	-histone acetyltransferase complex		-apicolateral plasma membrane -apical junction complex	-postsynaptic membrane -perinuclear region of cytoplasm -clathrin adaptor complex	-collagen -extracellular matrix

Table S6. GO terms enriched in HTC clusters by type content ($p < 0.01$)

Term	**Positive**	**Negative**	**Polar**	**Hydrophobic**	**Special**
GOBP	-ncRNA metabolic process -RNA processing -ribosome biogenesis -cell division -telomere organization -chromatin assembly or disassembly	-maintenance of fidelity during DNA-dependent DNA replication	-regulation of ARF protein signal transduction		-asymmetric protein localization
GOMF	-ATP-dependent helicase activity -purine NTP-dependent helicase activity -RNA binding -damaged DNA binding -cyclin-dependent protein kinase activity	-DNA secondary structure binding	-hyaluronic acid binding		
GOCC	-nucleoplasm -nuclear speck	-mismatch repair complex -endoplasmic reticulum part		-mitochondrial matrix -mitochondrial lumen	

Table S7. GO terms enriched in LTC clusters by type content ($p < 0.01$)

Term	**Positive**	**Negative**	**Polar**	**Hydrophobic**	**Special**
GOBP	-RNA processing -ncRNA metabolic process	-microtubule-based process	-DNA repair -response to inorganic substance	-coenzyme metabolic process -monovalent inorganic cation transport -coenzyme biosynthetic process	-response to insulin stimulus -response to peptide hormone stimulus
GOMF	-metallopeptidase activity -peptidase activity, acting on L-amino acid peptides	-DNA bending activity -sulfate transmembrane transporter activity	-histone methyltransferase activity -calcium ion binding -copper ion transmembrane transporter activity		-sodium channel activity -enzyme binding -diacylglycerol binding -kinase activator activity -hydrolase activity
GOCC	-intracellular organelle lumen -organelle lumen -membrane-enclosed lumen				-Golgi cisterna membrane

Table S8. Percentage of 3042 HTC (high type content low residue conservation) IDRs having at least 70% residues of the specified type in clusters of 10 or 20 residues.

Residue cluster type	10 residue cluster	20 residue cluster
Positive	1.18	0.10
Negative	2.66	0.46
Polar	11.24	1.15
Hydrophobic	1.55	0.26
Special	1.81	0.23
Total	**18.44**	**2.20**