

Annex: Patent Classification Code Co-occurrence Analysis for Animal Genetic Resources

This Annex describes the process used to identify technology clusters in the patent landscape for animal genetic resources.

Typically patent analysis proceeds by identifying individual patent classification codes (IPCs or CPCs) of relevance from an available dataset. The problem with this approach in the case of animal genetic resources is that it does not permit us to readily distinguish those classification codes that are only concerned with animal genetic resources for food and agriculture purposes from those involving animal genetic resources for other purposes or genetic materials with a non-animal or human origin.

To address this problem we applied a novel method. This method involves analysis and mapping of those classification codes that occur together (co-occur) within the patent records. The theory behind this is that patent documents in specific areas of technology (such as new breeds of animals or housing animals) should cluster together. These clusters and the linkages between them can be mapped using network mapping software to visualize distinct areas of technology and any relationships between areas of technology. We applied a statistical algorithm known as the Modularity Class algorithm to assist us with identifying clusters or communities of classification codes for our original dataset of 80,000 patent documents containing references to animal names (Latin and Common names). The size of the dataset was subsequently updated until October 2013 but the details of the structure remain the same. Further technical details of the methods are provided below.

The aim of co-occurrence analysis was to examine whether a distinct network could be identified for new breeds of animals within the patent data that could be distinguished from other areas of technology that involve animal genetic material. The answer to this question is yes.

Method

The original universe of patent publications consisting of 80,429 documents is taken and the associated patent classifications codes (International Patent Classification Codes and Cooperative Patent Classification Codes) are obtained for each record. For the purpose of the analysis the International Patent Classification and the Cooperative Patent Classification were treated as a combined classification scheme because the CPC is effectively simply a more detailed version of the IPC. In total, 41,167 patent

classification codes were associated with the 80,429 patent documents containing references to animals.

This data forms the basis of an undirected network graph for further analysis. The data is grouped initially by classification code noting the number of publications it references. Every classification code represents a node (vertex) within the network graph. The size of each node is based on the number of patent publications associated with the classification code. The next step involves identifying the links between codes. To do this the data is grouped by publication numbers each with a set of naturally ordered classification codes. Each set is processed in turn counting every time two codes occur together. By using this approach each association (per publication) between code A and B is recorded only once (because the graph is undirected). The end result of this is a set of edges formed by a source and target codes and a weight formed by the number of patent publications they appear together in. Edges are represented as lines linking nodes on the network map below.

The resultant graph consists of 41,167 nodes (classification codes) and 101,543 associated edges (links between nodes). We used open source Gephi graphing software to visualise and analyse the graph. To simplify the process of laying out (partitioning) the graph we reduced the overall number of nodes by filtering out classification nodes that referenced less than 10 patent publications. This filter is an appropriate amount suggested by the classification removal method discussed above that assesses the impact of removing codes on overall coverage. We remove the IPC codes starting with A61P (154 in all) because they are purely descriptive codes that describe types of medical disorders and their treatments, for example A61P31/00 is antiseptics. The end result is a graph with 8,003 nodes.

Modularity class is a method of partitioning graphs based on the communities of closely related nodes (Guillaume et al 2008). Nodes related by edges (enhanced by node weights) are clustered into small communities which are then iteratively aggregated into larger communities based on the strength of the links between nodes. We run the modularity class algorithm in Gephi with a resolution factor of one. This default value ensures the method is not biased towards small or large communities. The result is a network that is partitioned into over 3,000 communities.

The results are graphed and processed to breakdown each modularity class into its constituent classification codes. Publications are counted and the classification codes are ranked. This allows the dominant features of a certain cluster (modularity class) to

be identified. For example analysis shows that A61K38/00 dominates the modularity class for biotechnology which is an expected result as A61K38/00 references 14% of the total universe of publications by itself.

First Level Co-occurrence Analysis

The first step is to perform graph analysis at the top level. This consists of the 8,003 IPC and CPC classification codes that reference ten or more patent publications within our original (1976-2010) universal set (80,429 publications). Within Gephi the Modularity class algorithm is applied, the graph layout organised to inspect the different classes is demonstrated in Figure 1. The colours represent communities of classification codes around particular areas of technology. The labels refer to the number of the community (modularity class) with details provided in Table 1 below. Figure 2 illustrates the detail of the network focusing on biotechnology and new breeds of animals.

Figure 1: Gephi Network Map of Patent Classification (IPC/CPC) Co-occurrence

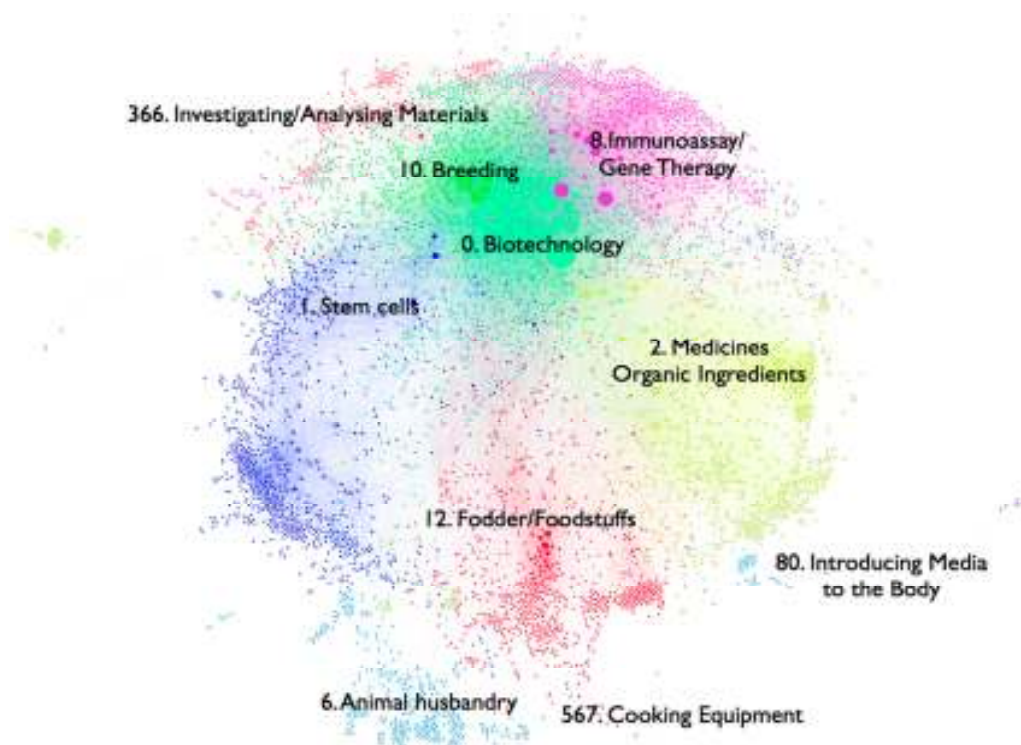


Figure 2. Core Network Detail

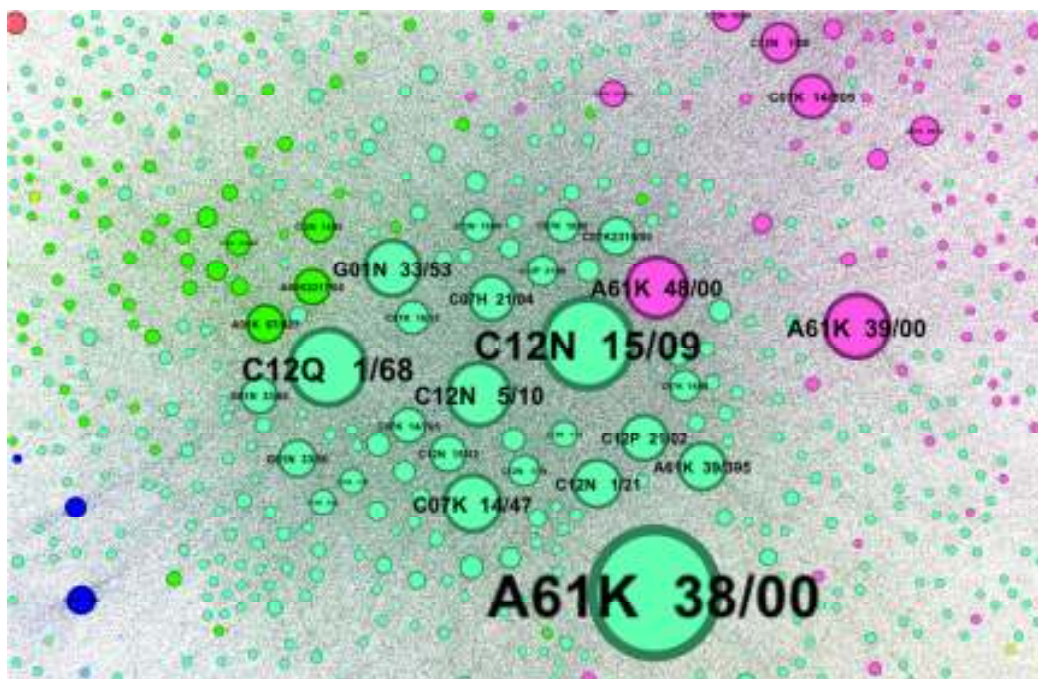


Table 1. Summary of Patent Classification (IPC/CPC) Modularity Classes

Summary	Key Classification Groups IPC and CPC	All Pubs	Latin Pubs	Class
Biotechnology.	Description in Table 2.	38506 47.78%	6519 86.63 %	0
Immunoassay/Gene therapy dominated by A61K39/00 and A61K48/00.	A01N43, A01N63, A01N65 (Biocides) A61K2039 (Medicines with antigens or antibodies: /522 Plant cells inactivated (killed); /5258 Virus like particles; /53 DNA vaccination; /552 veterinary vaccine; /55555 liposomes; /55561 adjuvants; /55566 emulsions) A61K35 (Medicines with: /76 viruses)	8653 23.0%	2277 30.2%	8

	<p>A61K38 (Peptide based medicines: /19 cytokines, lymphokines, interferons; /20 interleukins)</p> <p>A61K39 (Medicines containing antigens/antibodies: /0011 cancer antigens; /002 protozoan antigens; /02 bacterial antigens; /04 mycobacterium; /09 streptococcus)</p> <p>A61K48 (/00 Gene therapy)</p> <p>C07K14 (Peptides from viruses - Gastrins, Somatostatins, Melanotropin: /01 DNA Viruses; /02 hepatitis B; /03 rabies virus; /08 RNA viruses; /11 influenza; /15 retroviridae e.g. bovine leukaemia; /15 HIV-1, /18 togaviridae e.g. hepatitis C; /315 hafnia)</p> <p>C071K16 (Immuglobins: /08 against materials from viruses; /10 against material from RNA viruses; /12 against material from bacteria)</p> <p>C12N15 (Mutation/genetic engineering; /31 enterotoxins; /40 flaviviruses; /86 viral vectors; /861 adenoviral vectors; /863 poxviral vectors; /867 retroviral vectors; /869 herpesviral vectors)</p> <p>C12N2710 (dsDNA Viruses: /103343, /14143, /24143 virus (component) as vaccine; /16722 new viral proteins/genes)</p> <p>C12N2740 (Reverse transcribing RNA viruses: /13043 virus (component) as vaccine; /16122, /16222 new viral proteins/genes)</p> <p>C12N2760 (ssRNA Viruses - negative sense: /16122 new viral proteins/genes; /16134 virus (component) as vaccine)</p> <p>C12N7 (Virus vaccines: /01 modified by</p>			
--	--	--	--	--

	<p>introduction of foreign genetic material; /02 recovery or purification; /04 inactivation or attenuation)</p> <p>C12Q1 (Testing processes using microorganisms: /70 Geomicrobiological testing (petrol involving virus))</p> <p>C12R1 (Processes using micro organisms: /01 Bacteria; /92 Viruses)</p> <p>G01N2469 (Immunoassays for the detection of microorganisms; /20 detection of antibodies)</p> <p>G01N33 (Analysing materials: /Immunoelectrophoresis for micro organisms. e.g. protozoa, bacteria or viruses)</p>			
<p>Medicines with organic Ingredients dominated by A61K45/00 and A61K9/00.</p>	<p>A61D7 (Methods for introducing materials into bodies of animals)</p> <p>A61K31 (Active organic medicine ingredients: /16 amides; /19 carboxylic acid; /198 nitrofurans; /40 ascorbic acid; /44 pyridines; /47 quinoline; /505 pyrimidines; /519 barbituric acids; /56 steroids; /565 oestrone; /57 pregnane, progesterone; /70 D-galactosamine, ranimustine; /7048 levoglucosan; /715 polysaccharides)</p> <p>A61K33 (Heavy metals in medicines)</p> <p>A61K38 (Peptide based medicines: /04 gastrins; /21 interferons; /22 hormones; /23 calcitonins; /24 follicle-stimulating hormone; /28 Insulins;)</p> <p>A61K45 (Medicines with other active ingredients)</p> <p>A61K47 (Medicines by non-active ingredients /02 inorganic compounds; /10 alcohols; /12</p>	<p>18137</p> <p>22.36%</p>	<p>1599</p> <p>21.27%</p>	<p>2</p>

	<p>carboxylic acid; /18 amines; /26 carbohydrates /32 /34 macromolecular components; /36 polysaccharides; /42 proteins; /44 oils; /48 non-active ingredients)</p> <p>A61K9 (/00 Medicines by form: /0024 implanted; /06 ointments; /08 solutions; /107 emulsions; /127 liposomes; /14 powders; /16 granulates; /20 pills; /22 Release type; /48 capsules; /50 microcapsules)</p> <p>A61M31 (Remedies in cavities in body.)</p> <p>C07D401,405, 471 (Heterocyclic compounds)</p> <p>C07K1 (Peptide solutions)</p> <p>Y10S514 (Immune response affecting drug)</p>			
<p>Stem cells dominated by</p> <p>C12N5/00, A61K35/12 and C12N5/02.</p>	<p>A01N1 (Preservation of living parts)</p> <p>A61B10 (Diagnosis instruments)</p> <p>A61B17 -19 (Surgical instruments, devices methods)</p> <p>A61F2 (Prostheses implantable e.g. stents, artificial valves, bones)</p> <p>A61K35 (Preparations with materials from: /12 mammals or birds; /14 blood; /16 plasma; /28 bones)</p> <p>A61K38 (Peptide based medicines: /18 growth factors; /39 connective tissues e.g. collagen)</p> <p>A61L2 (Sterilization)</p> <p>A61L27 (Prosthesis material: /22 polypeptides; /24 collagen; /34,26 macromolecular; /3687 hollow organs e.g. bladder; /38 Animal cells,</p>	<p>12119</p> <p>14.94%</p>	<p>1060</p> <p>14.1%</p>	<p>1</p>

	<p>artificial skin; /54 active materials - A61L31/16)</p> <p>C07K14 (/51 Bone morphogenetic factor)</p> <p>C12N2500 (Cell culture medium)</p> <p>C12N2501 (Epidermal growth factor)</p> <p>C12N5 (/00 Undifferentiated human, animal or plant cells, e.g. cell lines, Tissues, Cultivation or maintenance thereof: /02 propagation; /0606 embryonic stem cells; /071 animal cells or tissues > Vertebrate cells or tissues, e.g. human cells or tissues; /0735 embryonic germ cells; /074 adult stems cells; /077 Mesenchymal cells e.g. bone cells)</p>			
Fodder/foodstuffs dominated by A23K1.	<p>A23C9 (Milk preparations)</p> <p>A23K1 (Animal feeding-stuffs)</p> <p>A23L1 (Foodstuffs)</p> <p>A23V2002 (Food compositions)</p> <p>A61K35 (Medical preparations with: /20 Milk; /54 Eggs; /74 Bacteria)</p> <p>A61K38 (Peptides based medicines: /01 hydrolysed proteins; /40 lactoferrins and ovo transferrins)</p>	1172 14.42%	616 8.2%	12
New Breeds of Animals dominated by A61K67/027, A01K2217 and C12N15/85.	Description in Table 5.	9963 12.28%	1151 15.3%	10
Animal husbandry.	A01K1 (Housing animals)	7972	190	6

	<p>A01K11 (Marking animals)</p> <p>A01K13 (Grooming animals)</p> <p>A01K15 (Taming animals)</p> <p>A01L29 (Other apparatus for animal husbandry)</p> <p>A01J5 (Pneumatics teats manipulation - milking)</p> <p>A01J7 (Milking machines)</p> <p>A22B5 (Slaughtering accessories)</p> <p>A22C21 (Poultry Processing)</p> <p>A61B5 (Veterinary ultrasound)</p>	9.83%	2.5%	
Investigating/Analysing biological material dominated by G01N33/543.	<p>B82Y5 (Nano-technology for materials)</p> <p>B82Y30 (Nanomedicines)</p> <p>C12M1 (Microbiology apparatus)</p> <p>G01N33 (Investigating materials. Various: antigens; antibodies; red blood cell count; immunosuppressants)</p>	5982 7.38%	404 5.37%	366
Plant resilience dominated by C12N15/82.	<p>A01H1 (/00 Processes modifying plant genotypes)</p> <p>A01H5 (Flowering Plants: /10 Seeds, e.g. gramineae leguminosae, brassicaceae)</p> <p>A01N63 (Biocides, repellants, regulators etc /02 from microorganisms and animal material)</p> <p>C07K14 (Peptides > 20 amino acids: /415 from plants)</p> <p>C12N15 (Mutation or genetic engineering: /29 genes encoding plant proteins; /82 for plant cells., e.g. plant artificial chromosomes (PACs);</p>	1768 2.2%	494 6.57%	16

	<p>/8243 metabolic engineering, e.g. nicotine, caffeine; /8257 for the production of primary gene products, e.g. pharmaceutical products, interferon; /8258 for the production of oral vaccines (antigens) or immunoglobulins)</p> <p>C12N5 (Undifferentiated human, animal or plant cells: /04 Plant cell tissues)</p>			
<p>Fertilizers and biofuel from agricultural byproducts dominated by C05F3/00.</p>	<p>B09B3 (Destroying solid waste)</p> <p>C02F2 (Galvanic waste from agriculture.)</p> <p>C02F3 (Biological water treatment)</p> <p>C05F3 (Fertilizers from animal waste)</p> <p>C05F17 (Composted fertilizers)</p> <p>Y02E50 (Grain bio-ethanol)</p>	<p>992</p> <p>1.22%</p>	<p>91</p> <p>1.21%</p>	<p>25</p>
<p>Computer security including trojan horses, guinea pig files etc dominated by G06F21/00.</p>	<p>G01F21 (Computing security)</p> <p>G06F12 (Computer memory protection)</p> <p>H04L29 (Communication control)</p>	<p>737</p> <p>0.9%</p>	<p>4</p> <p>0.05%</p>	<p>78</p>
<p>Enzymes used in detergents dominated by C12P41/00.</p>	<p>C12P7 (Carboxyl)</p> <p>C12P41(Enzyme using processes to separate optical isomers from racemic mixture)</p> <p>C11D3 (Detergent ingredients containing enzymes)</p>	<p>832</p> <p>1.03%</p>	<p>113</p> <p>1.5%</p>	<p>273</p>
<p>Cleaning/repairing pipes - Various mentions of pigs such as pipeline</p>	<p>B08B9 (Removing Blockages)</p> <p>F16L55 (Pipe maintenance equip.)</p>	<p>561</p> <p>0.69%</p>	<p>0</p> <p>0%</p>	<p>408</p>

pig or dispersal pig dominated by B08B9/04 and B08B9/055.				
Semi-permeable membranes and sorbents.	<p>A61M1 (Medical Pumping devices: /16 Dialysis systems)</p> <p>B01D15 (Treatment of liquids with solid sorbents: /08 chromatography)</p> <p>B01D16 (Separation with semi permeable membranes e.g. dialysis: /14 ultrafiltration)</p> <p>B01D67 (/00 Manufacturing semi-permeable membranes)</p> <p>B01D69 (Semi-permeable membranes: /02 storing or preservation; /08 hollow fibre membranes)</p> <p>B01J20 (Solid sorbent compositions: /28 Synthetic macromolecular compounds; /281 specially adapted for preparative chromatography; /32 Impregnating or coating; /54 Sorbents specially adapted for analytical or investigative chromatography)</p> <p>G01N30 (Investigating or analysing materials by separation into components using adsorption or ion-exchange, e.g. chromatography: /483 solid sorbents; /88 Integrated analysis systems)</p>	505 0.62%	22 0.29%	361
Veterinary Instruments, preparations and methods for administering	<p>A61M11 (Sprays)</p> <p>A61M15 (Inhalators)</p> <p>A61M16 (Respiratory)</p>	314 0.39%	4 0.05%	80

medicine into animals.	A61K9 (Medicine form)			
Tripeptide sweeteners.	C07K5 (Peptides with 4 amino acids: /072 Tripeptides; /0806, /1008 Aspartame Gly,Ala; /0808, /101 Val, LLe, Leu; /081, /1013 Aspartame Cys, Ser; /0812, /1016 Aspartame and aromatic; /0819, /1021 first amino acid being acidic; /0823 Aspartame and pro-amino acid; /083 Tripeptides, Gly, Ala, /087 Phe, Tyr; /09 Asp, Glu, Arg; /093 Asp, Glu, Arg; /097 Pro, His, Trp; /08, /10 Tetrapeptides; /1008 Aspartame Gly, Ala; /101 Aspartame Val, LLe, Leu; /1024 Aspartame heterocyclic; /103 Tetrapeptides Gly, Ala; /107 Phe, Tyr; /11 Lys, Arg; /113 Asp, Glu, Asn; /117 Pro, His, Trp)	294 0.36%	14 0.1%	9
Cooking equipment dominated by A47J37/04.	A47J37 (Cooking equipment: /04 spits; /041 scale prevention; /044 scale prevention inclined; /06 roasters/grills; /07 barbeques) A47J39 (/00 Heat insulated warming chambers) A47J43 (Implements to prepare/hold food: /18 clamping devices for fowl, venison or other meat, vegetables)	368 0.45%	0 0%	567
Saddles.	B68C1 (Saddling equipment: /002 for cleaning; /02 saddles; /025 saddle trees; /04 adjustable saddles; /08 side saddles; /14 straps for saddles; /16 fastening stirrups to saddles)	288 0.36%	0 0%	23
Saddles upholstery.	B68C1 (Saddling equipment: /12 Bottom pads for saddles)	101 0.12%	0 0%	4

Decoys.	A01M31 (Hunting appliances: /06 Decoys)	198	0	3
		0.24%	0%	
Whistles.	A01M31 (Hunting appliances: /004 game callers)	150	1	5
	A63H5 (/00 Musical or noise-producing devices for additional toy effects other than acoustical).	0.18%	0.01%	

Biotechnology Co-occurrence Analysis

The first level analysis presents a modularity class that we have labelled biotechnology that consists of 1,100 full IPC and CPC classification symbols. This level of analysis is unsatisfactory so we take a snapshot of all nodes within Biotechnology and perform Modularity class analysis again to identify the codes within biotechnology. The top ranking results are presented in Figure 3 and Table 2.

Figure 3. Network for Biotechnology

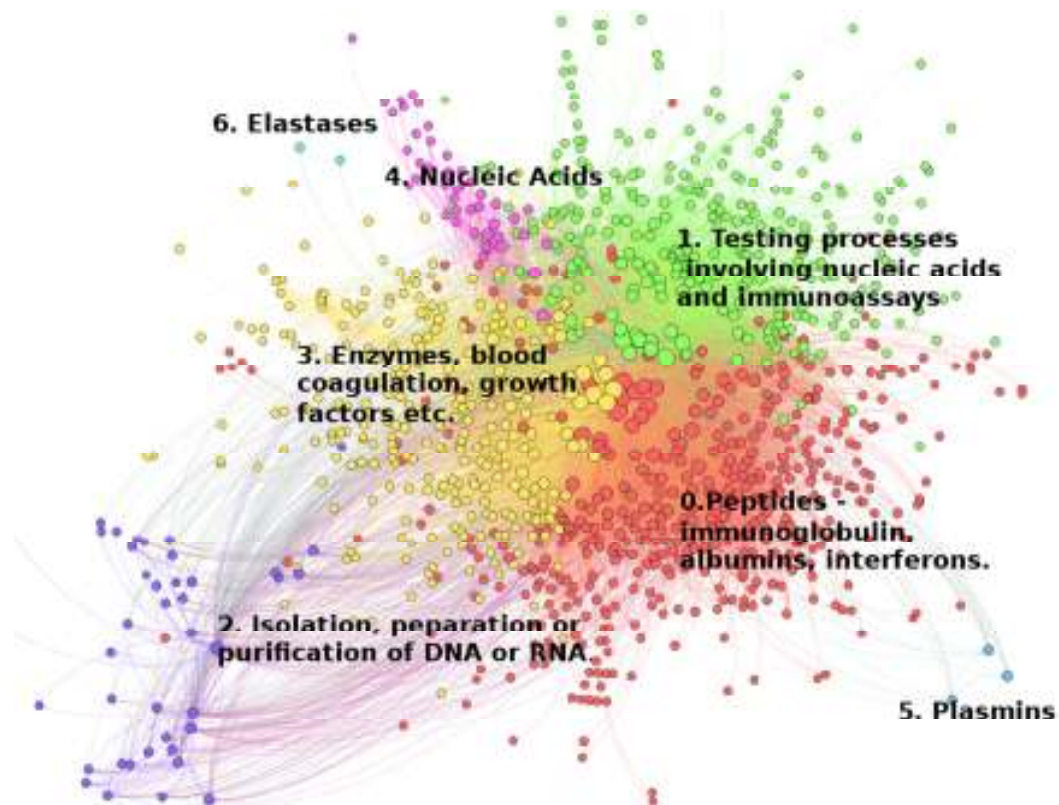


Table 2 - Biotechnology Modularity Classes Top Level Breakout

Summary	Key Classification Groups IPC and CPC	All Pubs	Latin Pubs	Class
Peptides - immunoglobulin, albumins, interferons dominated by A61K38/00, C07K14/47 and A61K39/395.	Description in Table 3.	25190 31.06 %	4597 61.16%	0
Testing processes involving nucleic acids and immunoassay dominated by C12Q1/68 and G01N33/53.	<p>A61K31 (Medicines with organic active ingredients: /7088 having three or more nucleosides or nucleotides; /7105 natural ribonucleic acids)</p> <p>C07H21 (Nucleic Acids: /02 heterocyclic radicals with oxygen or sulphur as ring hetero atoms)</p> <p>C12M1 (/00 enzymology or microbiology)</p> <p>C12N5 (Undifferentiated human, animal or plant cells, e.g. cell lines; Tissues; Cultivation or maintenance thereof: /06 Animal cells or tissues)</p> <p>C12P19 (Compounds containing saccharide radicals: /34 polynucleotides, e.g. nucleic acids)</p> <p>C12Q1 (/00, /02 Testing processes involving enzymes or microorganisms e.g. colony counters: /04 Determining presence or kind of</p>	18014 22.2%	3637 48.4%	1

	<p>microorganism; /018 antimicrobial activity of material; /25 involving certain enzymes; /26 methods of physically isolating intact microorganism involving oxidoreductase; /34 involving hydrolase; /37 involving peptidase; /42 involving phosphatase; /44 involving esterase; /48 involving transferase; /485 involving kinase; /527 involving lyase; /66 Geomicrobiological testing e.g. for petroleum involving luciferase; /68 involving nucleic acids; /6806 PCR assay; /6809 sequence identification involving differential detection; /686 polymerase chain reaction; /6881 e.g. Q-beta replicase for tissue and cell typing, e.g. HLA probes; /6886 Q-Beta replicase for cancer; /689 for bacteria; /6897 Q-beta replicas involving reporter genes operably linked to promoters)</p> <p>C12Q2600 (Oligonucleotides: /136 screening for pharmacological compounds; /156 polymorphic or mutational markers; /158 expression markers)</p> <p>C40B30 (Methods of screening libraries: /04 antibody-antigen binding, receptor-ligand binding)</p> <p>G01N21 (Optical investigation: /78 Bioluminescence producing a change of colour)</p> <p>G01N2500 (Screening for compounds of potential therapeutic value: /02 effect of C on interaction between molecules A and B; /04 effect of C on molecule A; /10 involving cells)</p>			
--	--	--	--	--

	<p>G01N2800 (Disease detection: /2821 Alzheimer; /2828 Prion diseases; /52 predicting or monitoring response to treatment)</p> <p>G01N33 (investigating materials: /15 medicine; /48 biological material e.g. blood or urine; /483 physical analysis of biological material; /50 chemical analysis of biological material; /5011 partitioning blood components, testing antineoplastic activity; /5017 neoplastic activity; /5041 involving analysis of members of signalling pathways; /5088 mitochondria of vertebrates; /5091 mitochondria for testing pathological state of organism; /53 Immunoassay; /5308 improving reaction conditions; /533 promotion of specific binding with fluorescent label; /542 fluorescence quenching; /554 bacteria, yeast cells; /564 immunoelectrophoresis for autoimmune disease; /566 immunoelectrophoresis as ligand binding reagent; /567 immunoelectrophoresis using organ as binding agent; /56972 white blood cells; /573 immunoelectrophoresis for enzymes; /574 Immunoelectrophoresis for cancer; /57415 cytomegalovirus, Epstein-Barr virus of breast; /57419 of colon; /57449 of ovaries; /5748 involving oncogenic proteins; /57496 involving intracellular compounds; /58 Immunoelectrophoresis involving labelled substances; /582 Hepatitis A with fluorescent label; /68 immunoelectrophoresis involving proteins, peptides or amino acids; /6842 membrane proteins; /6872 Intracellular protein regulatory factors; /689</p>			
--	---	--	--	--

	<p>nucleoproteins related to pregnancy; /6896 Alzheimer's disease; /72 immunoelectrophoresis involving blood pigments; /74 involving hormones; /743 steroid hormones; /92 thyroid gland hormones)</p> <p>G06F17 (Information retrieval: /30 Databases)</p> <p>G06F19 (Specific applications of digital computing: /18 totalisers for functional genomics or proteomics, genotype-phenotype associations; /22 totalisers for sequence comparison involving nucleotides or amino acids)</p>			
<p>Enzymes, blood coagulation, growth factors dominated by C12N15/09 and C12N5/10.</p>	<p>Broken out in Figure 5 and Table 4.</p>	<p>20852 25.71 %</p>	<p>4061 54.03%</p>	<p>3</p>
<p>Nucleic Acids dominated by C12N15/11 and C12N15/113.</p>	<p>A61K31 (Medicines with organic active ingredients: /713 double stranded nucleic acids or oligonucleotides)</p> <p>C12N15 (Mutation or genetic engineering: /11 DNA or RNA fragments; /111 general methods applicable to biologically active non-coding nucleic acids; /1137 against enzymes; /1138 against receptors or cell surface proteins; /113 non-coding nucleic acids modulating the expression of genes, e.g. antisense oligonucleotides)</p> <p>C12N2310 (Structure or type of nucleic acid:</p>	<p>1404 1.73%</p>	<p>346 4.6%</p>	<p>4</p>

	/111 indexing codes of methodologies for biologically active non-coding nucleic acids spanning whole gene; /315 Phosphorothioates; /321 2'-O-R Modification)			
Isolation, preparation or purification of DNA or RNA dominated by C12N15/10.	C07K14 (Peptides having more than 20 amino acids, Gastrins, Somatostatins and Melanotropins: /805 Haemoglobins) C12N15 (Mutation or genetic engineering: /10 isolation, preparation or purification of DNA or RNA; /1037 screening libraries on surface of microorganisms, e.g. phage display, e. coli display; /1096 general methods of preparing gene libraries, cDNA synthesis; subtracted cDNA library construction) C40B40 (Libraries per se: /02 libraries contained in or displayed by microorganisms. e.g. bacteria or animal cells) G01N33 (Analysing materials: /6845 methods of identifying protein-protein interactions in protein mixtures)	1665 2.05%	251 3.34%	2
Plasmins.	C12N9 (Enzymes, e.g. ligases Proenzymes; Compositions thereof: /6435 Plasmin, i.e. fibrinolysin; /68 Plasmin) C12Y304 (Hydrolases acting on peptide bonds i.e. peptidases: /21007 Plasmin)	84 0.1%	5 0.06%	5
Elastases.	C12N9 (Enzymes, e.g. ligases Proenzymes; Compositions thereof: /6448 Elastases, e.g. pancreatic elastase; /66 Elastase)	21 0.02%	0 0%	6

Peptide Co-occurrence Analysis

Figure 4. Network for Peptides

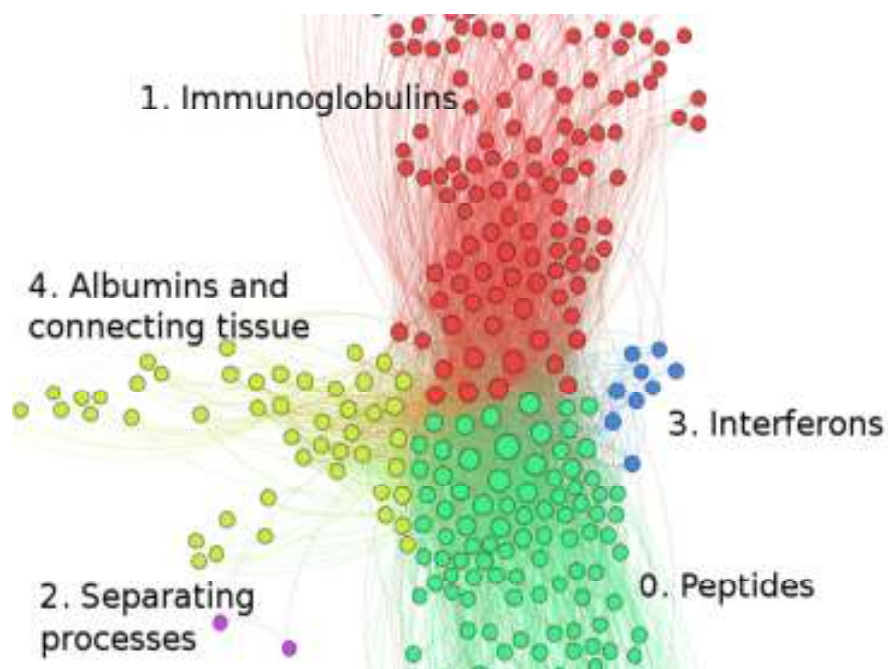


Table 3 - Peptide Modularity Class Analysis

Summary	Key Classification Groups IPC and CPC	All Pubs	Latin Pubs	Class
Peptides dominated by A61K38/00.	<p>A01N37 (Biocides: /18 Aromatic or aliphatic carboxylic acids)</p> <p>A61K38 (Medicines with peptides: /00 peptides; /02 undefined number of amino acids; /08 5-11 amino acids; /10 - 12 to 20 amino acids; /12 cyclic peptides; /16 gastrins, somatostatins, melanotropins; /17 gastrins etc from humans; /42 haemoglobins)</p> <p>C07K1 (Preparation of Peptides: /113 without changing primary structure; /14 extraction; /16 purification by</p>	20510 25.29 %	4077 54.24%	0

	<p>chromatography; /18 Ion-exchange chromatography; /20 partition interaction chromatography; /22 affinity chromatography)</p> <p>C07K14 (Peptides: /435 Thaumatin from animals or humans; /46 Eimeria from vertebrates; /465 from birds; /47 from mammals; /4703 inhibitors; Alzheimer's disease; /4713 Autoimmune diseases (insulin - dependent diabetes); /4747 Apoptosis related proteins; /4748 Tumour specific antigens; /475 Growth factors; /495 transforming growth factor; /52 cytokines, lymphokines, interferons; /54 Interleukins; /55 IL-2; /55 Interferons; /56 IFN-alpha; /575 Hormones; /60 Somatoliberin; /635 Parathormone; /705 receptors; /70571 nuclear receptors; /71 molecules with "CD" designation; /715 Receptors, Cell surface antigens; /72 Cell surface determinants for hormones; /723 G protein coupled receptor; /82 translation products from oncogenes)</p> <p>C07K16 (Immunoglobulins, e.g. monoclonal or polyclonal antibodies: /26 interferons against hormones)</p> <p>C07K17 (/00 Carrier bound or immobilized peptides)</p> <p>C07K19 (/00 Hybrid peptides)</p> <p>C07K2 (/00 Peptides with undefined number of amino acids)</p> <p>C07K2299 (/00 Coordinates from 3D structures of peptides)</p>			
--	--	--	--	--

	<p>C07K5 (/00 Peptides with 4 amino acids)</p> <p>C07K7 (/00 Peptides having 5 to 20 amino acids in a fully define sequence: /06 5 to 11 amino acids; /08 12 to 20 amino acids; /23 Luteinising hormone-releasing hormone)</p> <p>C12N15 (Mutation or genetic engineering: /12 Immunoglobulins.genes encoding animal proteins; /19 Interferons, lymphokines; cytokines)</p> <p>C12N2799 (Uses of viruses: /026 for redistribution where vector is from baculovirus)</p> <p>C12R1 (Microorganisms; /91 Trichoderma of viruses or cell lines)A61K39 (Medicines with antigens or antibodies: /395 Antibodies, Immunoglobulins; Immune serum)</p>			
<p>Immunoglobulins dominated by A61K39/395.</p>	<p>A61K39 (Medicines with antigens or antibodies: /395 Antibodies, Immunoglobulins; Immune serum)</p> <p>A61K47 (Medicines by non-active ingredients: /48284 Redox delivery systems)</p> <p>A61K49 (Test preparations /00 in vivo)</p> <p>C07K16 (Immunoglobulins: /06 from serum; /065 purification; /18 against material from animals or humans; /20 from protozoa; /22 plasmodium against growth factors; /24 against cytokines, lymphokines or interferons; /28 against receptors; /2863 interferons against receptors for growth factors; /30 from tumour cells; /32 reproductive system against translation</p>	<p>10186</p> <p>12.56 %</p>	<p>2039</p> <p>27.13%</p>	<p>1</p>

	<p>products of oncogenes; /42 anti-idiotypic antibodies; /44 reproductive system against material not provided elsewhere; /46 hybrid immunoglobulins with peptides)</p> <p>C07K2316 (Immunoglobulins specific features: /52 Constant or Fc region; /96 antibodies with antagonist activity)</p> <p>C07K2317 (Immunoglobulins specific features: /21 specific culture conditions from primates; /22 from camelid; /24 from different specie; /31 multispecific; /34 identification of a linear epitope; /55 fab; /565 complementary determining region; /569 single domain ; /622 single chain antibody; /732 antibody-dependant cellular cytotoxicity; /92 affinity, association rate, dissociation rate)</p> <p>C07K2319 (Fusion polypeptide: /30 Non-immunoglobulin-derived peptide)</p> <p>C12N15 (Mutation or genetic engineering: /02 hybrid cell by fusion; /13 Immunoglobulins)</p> <p>C12N5 (Undifferentiated human, animal or plant cells: /20 Murine cells, e.g. mouse cells one of the fusion partners being a B lymphocyte)</p> <p>C12P21 (Preparation of peptides or proteins: /08 Monoclonal antibodies)</p> <p>C12R1 (Microorganisms; /91 Trichoderma of viruses or cell lines)</p> <p>G01N33 (Analysing materials: /577</p>			
--	---	--	--	--

	Hepatitis A involving monoclonal antibodies)			
Albumins and connecting tissue dominated by Y10S435/975 and C07K14/78.	<p>A61K38 (Medicines with peptides: /38 albumins)</p> <p>C07K14 (Peptides: /76 Albumins; /765 Serum albumin e.g. HSA; /78 connective tissue peptides)</p> <p>G01N33 (Analysing materials: /531 immunochemical test materials; /6887 nucleoproteins from muscle, cartilage or connecting tissue; /94 thyroid gland hormones)</p> <p>Y10S435 (chem : Molecular biology and microbiology: /81 packaged kit; /975 Kit)</p> <p>Y10S436 (chem: analytical and immunological testing: /804 radioisotope; /813 cancer)</p> <p>Y10S530 (chem: peptides/proteins: /806 antigenic peptides; /807 hapten conjugated with peptide)</p>	3597 4.44%	281 3.74%	4
Interferons dominated by C07K14/555 and C07K14/56.	C07K14 (Peptides: /555 Interferons (IFN); /56 IFN-alpha)	390 0.48%	92 1.23%	3
Separating processes.	B01D15 (Separating processes involving the treatment of liquids with solid sorbents: /18 selective absorption e.g. chromatography relating to flow patterns; /1807 Flash chromatography using counter-currents, e.g. fluidised beds)	11 0.01%	0 0%	2

Enzymes, blood coagulation, growth factors Co-occurrence Analysis

Figure 5. Network for Enzymes, blood coagulation and Growth Factors

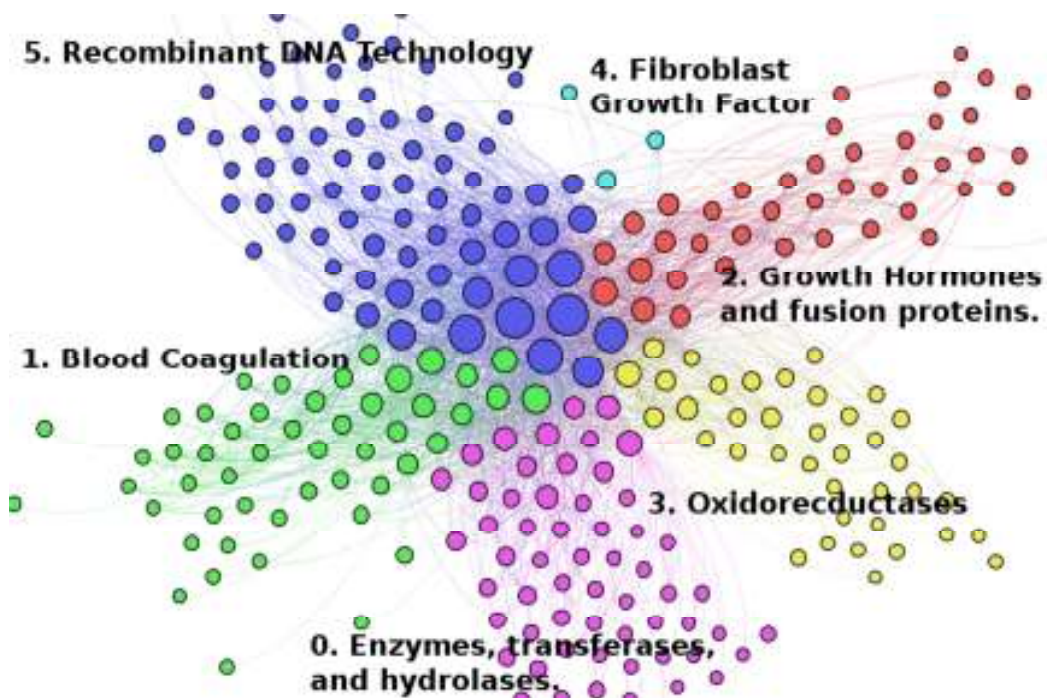


Table 4. Enzymes, blood coagulation and Growth Factors Mod. Class Analysis

Summary	Key Classification Groups IPC and CPC	All Pubs	Latin Pubs	Class
Recombinant DNA technology with vectors dominated by C12N15/09 and C12N5/10.	A61K31 (Medicines with organic active ingredients: /711 Natural deoxyribonucleic acids, i.e. containing only 2'-deoxyriboses attached to adenine, guanine, cytosine or thymine and having 3'-5' phosphodiester links) C07H21 (Nucleic Acids: /04 Heterocyclic radicals containing oxygen or sulfur as ring hetero atoms with deoxyribose as saccharide radical)	16,290 20.09 %	3,321 44.19%	5

	<p>C07K14 (Peptides having more than 20 amino acids, Gastrins, Somatostatins and Melanotropic: /195 rubella virus from bacteria; /41 delta-endotoxin from lichens; /535 Granulocyte; /59 stimulating hormones e.g. Follicle-stimulating hormone (FSH); /65 Somatomedins, e.g. IGF-1; /70567 Nuclear receptors)</p> <p>C12N1 (Microorganisms, e.g. protozoa: /15 fungi (culture of mushrooms); /16 yeasts; /18 baker's yeast; /19 brewer's yeast with introduced foreign genetic material; /20 Bacteria; /21 Bacteria with introduced foreign genetic material)</p> <p>C12N15 (Mutation or genetic engineering: /09 Recombinant DNA-technology; /16 hormones; /63 vectors; /64 general vector methods; /70 Vectors or expression systems specially adapted for E. coli; /74 specially adapted for lactobacillus, micromonospora; /79 specially adapted for eukaryotic hosts)</p> <p>C12N5 (Undifferentiated human, animal or plant cells, e.g. cell lines; Tissues; Cultivation or maintenance thereof: /10 virus-transformed cells; /16 animals cells)</p> <p>C12N9 (Enzymes e.g. ligases; Proenzymes; Compositions thereof : /02 luciferase; /04 glucose oxidase; /06 luciferase acting on nitrogen containing compounds as donors; /10 transferases; /12 transferases transferring phosphorus containing groups; /14 Hydrolases; /16 hydrolases acting on ester bonds; /18 Carboxylic ester hydrolases; /22 ribonucleases; /48 thromboplastin, leucine aminopeptidase; /50 Proteinases; /52 derived from bacteria; /64 bacillus derived from animal tissue, e.g. renin; /88 lyases; /90 isomerases; /96 forming</p>			
--	---	--	--	--

	<p>enzyme conjugates; /99 Enzyme inactivation by chemical treatment)</p> <p>C12P1 (/00 Compositions by using microorganisms or enzymes)</p> <p>C12P21 (Peptide/protein preparation: /02 having known sequence of two or more amino acids, e.g. glutathione; /04 cyclic or bridged peptides or polypeptides; /06 produced by hydrolysis of a peptide bond, e.g. hydrolysate products)</p> <p>C12R (Microorganisms: /19 Escherichia coli; /645 fung; /865 saccharomyces cerevisiae)</p>			
<p>Enzymes based medicines - transferases, hydrolases etc dominated by C07K16/40 and C12N15/00.</p>	<p>A61K38 (Medicines with peptides: /45 Transferases)</p> <p>C07K16 (Immunoglobulins: /40 Reproductive system, e.g. ovaria, uterus, testes, prostate against enzymes)</p> <p>C12N15 (/00 Mutation or genetic engineering: /52 genes encoding for enzymes; /54 transferases; /55 hydrolases)</p> <p>C12N9 (Enzymes e.g. ligases; Proenzymes; Compositions thereof : /10 transferases; /12 transferases transferring phosphorus containing groups; /14 Hydrolases; /16 hydrolases acting on ester bonds; /18 Carboxylic ester hydrolases; /22 ribonucleases; /88 lyases; /90 isomerases; /93 Ligases)</p>	<p>4,267</p> <p>5.7%</p>	<p>1,240</p> <p>16.5%</p>	<p>0</p>
<p>Blood Coagulation dominated by C12N9/64.</p>	<p>A61K38 (Medicines with peptides: /36 blood coagulation; /43 enzymes, proenzymes; /46 Hydrolases; /48 Hydrolases acting on peptide bonds; /55 protease inhibitors)</p>	<p>3,220</p> <p>4.0%</p>	<p>651</p> <p>8.66%</p>	<p>1</p>

	<p>C07K14 (Peptides having more than 20 amino acids, Gastrins, Somatostatins and Melanotropic: /745 blood coagulation factors; /755 Factors VIII; /775 Apo Lipopeptides; /81 protease inhibitors)</p> <p>C12N15 (Mutation or genetic engineering: /57 hydrolases acting on peptide bonds)</p> <p>C12N9 (Enzymes e.g. ligases; Proenzymes; Compositions thereof : /50 Proteinases; /52 derived from bacteria; /64 bacillus derived from animal tissue, e.g. renin; /99 Enzyme inactivation by chemical treatment)</p>			
<p>Growth hormones and fusion proteins dominated by C12N15/62, C07K2319/02 and C07K14/61.</p>	<p>C07K14 (Peptides having more than 20 amino acids, Gastrins, Somatostatins and Melanotropic: /61 growth hormone; /62 insulins)</p> <p>C07K16 (Immunoglobulins: /40 reproductive system against enzymes)</p> <p>C07K2319 (Fusion polypeptide: /02 Antigen-binding scaffold molecules; /50 Non-immunoglobulin-derived peptide, attached thereto containing protease site; /61 attached thereto containing an enzyme fusion for detection; /75 thrombopoietin, NPY and other peptide hormones)</p> <p>C12N15 (Mutation or genetic engineering: /18 growth hormones; /62 DNA sequences coding for fusion proteins; /81 for eukaryotic hosts for yeasts)</p> <p>Y01S930 (Peptide or protein sequence: /12 Growth hormone, growth factor other than t-cell or b-cell growth factor, and growth hormone releasing factor)</p>	<p>2,921 3.6%</p>	<p>477 6.35%</p>	<p>2</p>
<p>Oxidoreductases</p>	<p>A61K38 ((Medicines with peptides: /44</p>	<p>1,555</p>	<p>329</p>	<p>3</p>

<p>dominated by C12N9/02</p>	<p>Oxidoreductases)</p> <p>C07K14 (Peptides having more than 20 amino acids, Gastrins, Somatostatins and Melanotropic: /775 Apolipoproteins)</p> <p>C12N15 (Mutation or genetic engineering: /53 oxidoreductases)</p> <p>C12N5 (Undifferentiated human, animal or plant cells, e.g. cell lines; Tissues; Cultivation or maintenance thereof: /10 virus-transformed cells; /16 animals cells)</p> <p>C12N9 (Enzymes e.g. ligases; Proenzymes; Compositions thereof: /0006 JN acting on CH-OH groups as donor; /02 Oxidoreductases (1.), e.g. luciferase; /04 glucose oxidase, lactate dehydrogenase; /06 Oxidoreductases (1.), e.g. luciferase acting on nitrogen containing compounds as donors; /96 Stabilising an enzyme by forming an adduct or a composition)</p>	<p>1.9%</p>	<p>4.38%</p>	
<p>Fibroblast Growth Factor.</p>	<p>C07K14 (Peptides having more than 20 amino acids, Gastrins, Somatostatins and Melanotropic: /50 Fibroblast growth factor (FGF); /501 Transforming growth factor (TGF) acidic FGF (aFGF); /503 Transforming growth factor (TGF) basic FGF (bFGF))</p>	<p>174</p>	<p>33 0.44%</p>	<p>4</p>

New Breeds of Animals Co-occurrence Analysis

The modularity class for new breeds of animals is further decomposed by capturing the 184 relevant classification codes as the graph nodes and recalculating the modularity. Experimental research demonstrates that a resolution parameter of 0.5 produces the appropriate amount of classes, 15 in total. The resolution parameter n is a factor applied to the modularity class algorithm that either promotes a higher number smaller communities ($n < 1$) or ($n > 1$) a smaller number of larger communities. Our approach ends with decomposed classes, as shown in Figure 6 and Table 5, are either dominated specific classification codes such as A01K67/027 dominates new breeds of vertebrates class.

Figure 6. Network for New Breeds of Animals

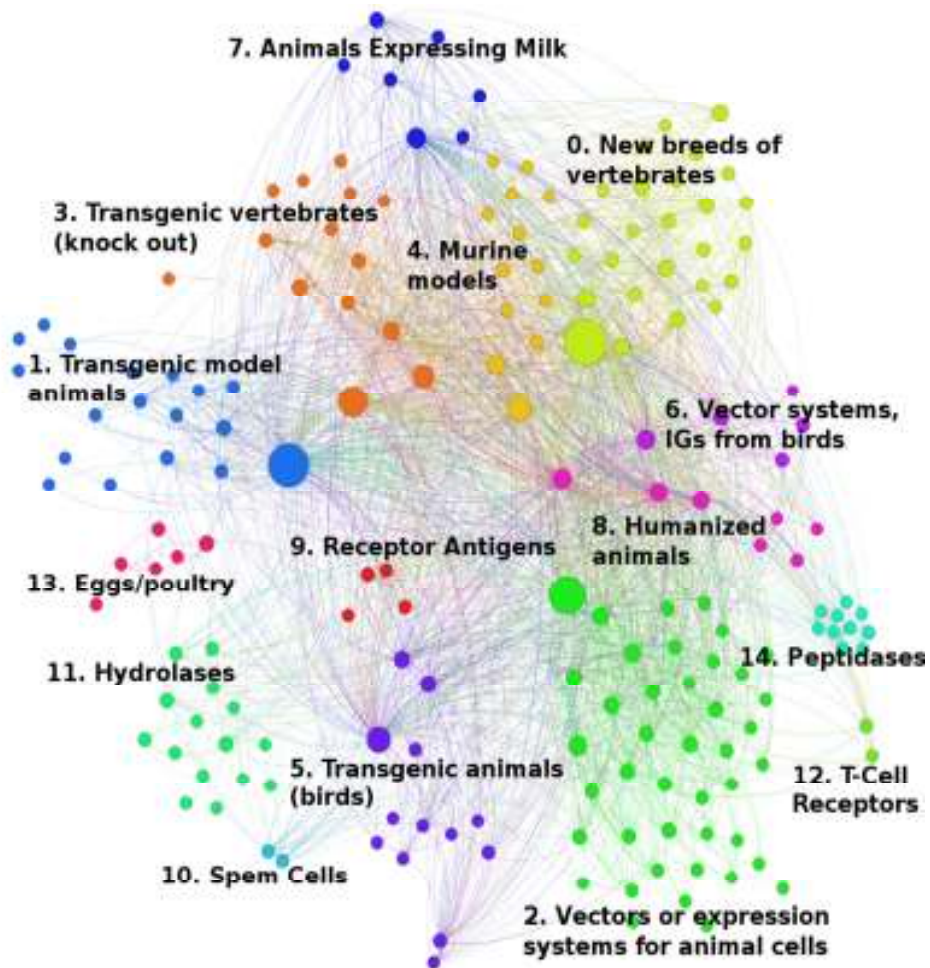


Table 5. New Breeds of Animals Modularity Class Analysis

Summary	Key Classification Groups IPC and CPC	All Pubs	Latin Pubs	Class
Breeding.	Overview.	9963 12.28%	1151 15.3%	10
New breeds of vertebrates dominated by A01K67/027.	<p>A01K2227 (Animals characterised by species: /101 Bovine; /102 Caprine; /103 Ovine; /106 Primate; /108 Swine)</p> <p>A01K67 (Rearing or breeding animals: /02 Breeding vertebrates; /027 New Breeds of vertebrates; /071 Chimeric animals, e.g. comprising exogenous cells; /0273 Cloned animals)</p> <p>C12N15 (Mutation or genetic engineering: /873 nuclear transfer, manipulation of totipotent cells or production of chimeric embryos; /877 Techniques for producing new mammalian cloned embryos; /8771 cloned bovine embryos; /8772 cloned caprine embryos; /8773 cloned ovine embryos; /8775 cloned murine embryos; /8778 cloned swine embryos)</p> <p>C12N2517 (Cells relating to new breeds of animals: /02 cells from transgenic animals; /04 cells produced using nuclear transfer; /10 conditioning of cells for in vitro fecondation or nuclear transfer)</p> <p>C12N5 (Undifferentiated human, animal or plant cells, e.g. cell lines, Tissues, Cultivation: /0604 Whole embryos; culture medium therefor; /0609 Vertebrate cells, production of embryos,</p>	3988	664	0

	nuclear transfer (A01K67/027) > oocytes, oogonia, fertilised oocytes; /075 oocytes, oogonia) G01N33 (/00 analysing materials)			
Transgenic vertebrates (knock out) dominated by C12N15/8509 and A01K2217/075.	A01K2217 (Genetically modified animals: /072 Animals genetically altered by homologous recombination > maintaining or altering function, i.e. knock in; /075 inducing loss of function, i.e. knock out) A01K2267 (Animals characterised by purpose: /02 Animal zoo technically ameliorated; /025 Animal producing cells or organs for transplantation; /0325 Animal models for autoimmune diseases; /0337 models for infectious diseases; /0343 models for processes and diseases of the central nervous system, e.g. stress, learning, schizophrenia, pain, epilepsy; /0381 model for diseases of the hematopoietic system) A01K67 (Rearing or breeding animals: /0276 Genetically modified vertebrates, e.g. transgenic > Knockout animals) C12N15 (Mutation or genetic engineering: /8509 Recombinant DNA-technology > Introduction of foreign genetic material using vectors > for animal cells > for producing genetically modified animals, e.g. transgenic) C12N9 (Enzymes, Proenzymes, Compositions thereof: /01087 alpha-1,3-galactosyltransferase)	2292	506	3
Transgenic vertebrates (birds) dominated by	A01K2217 (Genetically modified animals: /052 Animals comprising random inserted nucleic acids (transgenic) > inducing gain of function; /15 animal model comprising tissue-specific	1943	326	5

A01K67/0275.	<p>expression system, e.g. tissue specific expression of transgene, of Cre recombinase)</p> <p>A01K2227 (Animals characterised by species: /30 Bird)</p> <p>A01K67 (Rearing or breeding animals: /0275 New breeds of vertebrates > Genetically modified vertebrates, e.g. transgenic)</p> <p>C07K14 (Peptides > 20 amino acids: /77 from animals > Albumins > Ovalbumin)</p> <p>C12N15 (Mutation or genetic engineering: /87 Introduction of foreign genetic material using processes not otherwise provided for, e.g. co-transformation)</p> <p>C12N2799 (Uses of viruses: /022 as vector > for the expression of a heterologous nucleic acid > vector derived from an adenovirus; /027 from a retrovirus)</p> <p>C12N2800 (Nucleic acids vectors: /20 Pseudochromosomes, minichromosome; /204 of bacterial origin, e.g. BAC; /206 of yeast origin, e.g. YAC, 2u; /208 of mammalian origin, e.g. minichromosome)</p>			
Humanized animals (knock-in) dominated by A01K2217/00, A01K2207/15 and A01K67/0278.	<p>A01K2207 (Modified animals: /15 humanized animals)</p> <p>A01K2217 (/00 Genetically modified animals)</p> <p>A01K2227 (Animals characterised by species: /107 mammal > Rabbit)</p> <p>A01K67 (Rearing or breeding animals: /0278 New breeds of vertebrates > Genetically modified vertebrates, e.g. transgenic > humanized</p>	316	108	8

	<p>animals, e.g. knockin)</p> <p>C07K14 (Peptides > 20 amino acids: /20 from animals, from humans > Blood coagulation or fibrinolysis factors > Fibrinogen)</p> <p>C017K2317 (Immunoglobulins specific features: /20 characterized by taxonomic origin)</p> <p>C12N2800 (Nucleic acids vectors: /80 Vectors containing sites for inducing double-stranded breaks, e.g. meganuclease restriction sites)</p>			
Sperm cells.	<p>C12N5 (Undifferentiated human, animal or plant cells, e.g. cell lines, Tissues, Cultivation or maintenance thereof: /061 Animal cells or tissues;, Human cells or tissues > Vertebrate cells > Germ cells, production of embryos, nuclear transfer > Sperm cells, spermatogonia; /076 Animal cells or tissues > Sperm cells, Spermatogonia.</p>	92	6	10
Transgenic model animals with random inserted nucleic acids dominated by A01K2217/05.	<p>A01K2217 (Genetically modified animals: /05 Animals comprising random inserted nucleic acids (transgenic); /30 Animal model comprising expression system for selective cell killing, e.g. toxins, enzyme dependent prodrug therapy using ganciclovir)</p> <p>A01K2227 (Animals characterised by species: /40 Fish; /50 Amphibians; /703 Worms e.g. Caenorhabditis elegans; /706 Insects, e.g. Drosophila melanogaster)</p> <p>A01K2267 (Animals characterised by purpose: /0312 animal model for Alzheimer's disease; /0318 animal model for neurodegenerative disease, e.g. non- Alzheimer's; /0368 animal model for</p>	2994	552	1

	<p>inflammation)</p> <p>A01K67 (Rearing or breeding animals: /033 New breeds of invertebrates, non-chemical sterilization of invertebrates; /0333 Genetically modified invertebrates, e.g. transgenic, polyploid; /0336 Genetically modified Nematodes, e.g. Caenorhabditis elegans; /0338 Genetically modified Crustaceans; /0339 Genetically modified insects, e.g. Drosophila melanogaster, medfly)</p> <p>C12N2799 (Uses of viruses: /021 as vector for the expression of a heterologous nucleic acid)</p> <p>C12N9 (Enzymes, Proenzymes, Compositions thereof: /1048 Glycosyltransferases)</p>			
<p>Vector or expression systems for animal cells dominated by C12N15/85.</p>	<p>A61K47 (Medicines with non-active ingredients, e.g. carriers: /48776 cells, cell fragments, viruses, ghosts, red blood cells or viral vectors)</p> <p>A61K48 (Gene therapy: /0066 Manipulation of the nucleic acid to modify its expression pattern)</p> <p>C07K14 (Peptides > 20 amino acids: /505 from animals, from humans, growth factors Erythropoietin (EPO); /605 hormones, Glucagons)</p> <p>C07K16 (Immunoglobulins, e.g. monoclonal or polyclonal antibodies: /02 Metallothioneins from eggs)</p> <p>C07K2319 (Fusion polypeptide: /09 Antigen-binding scaffold molecules)</p> <p>C12N15 (Mutation or genetic engineering notably recombinant DNA technology, introducing foreign genetic material using vectors: /635 tetR inducible by tetracycline; /65, enzymes used as markers /85</p>	4092	472	2

	<p>Vectors or expression systems specially adapted for eukaryotic hosts i.e. for animal cells; /67 general methods for enhancing expression; /8213 Targeted insertion of genes into the plant genome by homologous recombination; /90 Stable introduction of foreign DNA into chromosome; /907 Swine embryos in mammalian cells)</p> <p>C12N2510 (Genetically modified cells: /02 cells for production)</p> <p>C12N2800 (Nucleic acids vectors: /107 for redistribution for mammalian; /108 for episomal vectors; /60 vectors containing traps for, e.g. exons, promoters; /90 Vectors containing a transposable element)</p> <p>C12N2830 (Vector systems having a special element relevant for transcription: /003 C. albicans tet inducible; /006 C. albicans tet repressible; /15 chimeric enhancer/promoter combination; /30 being an enhancer not forming part of the promoter region; /36 being a transcription termination element; /38 being a stuffer; /40 being an insulator; /42 being an intron or intervening sequence for splicing and/or stability of RNA; /46 elements influencing chromatin structure; /55 C. albicans from bacteria; /85 C. albicans mammalian; /90 C. albicans avian)</p> <p>C12N2840 (Vectors comprising a special translation-regulating system: /105 C. albicans enhancing translation)</p>			
Hydrolases dominated by C12N9/24 and	<p>A61K38 (Peptide based medicine: /47 cellulases; lactases)</p> <p>C07K14 (Peptides > 20 amino acids: /79 from</p>	675	89	11

<p>A61K38/47 and C07K14/79.</p>	<p>animals, from humans > transferrins, e.g. lactoferrins, ovotransferrin)</p> <p>C12N15 (Mutation or genetic engineering; /56 Genes encoding for enzymes or proenzymes > Hydrolases, e.g. amylase, galactosidase, lysozyme)</p> <p>C12N9 (Enzymes, Proenzymes, Compositions thereof: /24 Hydrolases acting on glycosyl compounds; /2402 hydrolysing O- and S- glycosyl compounds; /2408, /2434 acting on alpha -1,4- glucosidic bonds; /2462 Lysozyme; /26 hyaluronidase, invertase, amylase ; /36 lysozyme)</p> <p>C12Y302 (Enzymes > hydrolases: /01014 Chitinase; /01018 trans-sialidase; /01031 Beta- glucuronidase; /01052 Beta-N- acetylhexosaminidase; /01166 Heparanase)</p>			
<p>Murine models for test or diseases, dominated by A01K2227/105 and A01K2267/03.</p>	<p>A01K2217 (Genetically modified animals: /20 Animal model comprising regulated expression system)</p> <p>A01K2227 (Animals characterised by species: /105 Mammal > Murine)</p> <p>A01K2267 (Animals characterised by purpose: /03 model for test or diseases; /306 model for genetic diseases; /0331 model for proliferative diseases; /035 model for multifactorial diseases; /0362 model for obesity, type-2 diabetes; /0375 model for cardiovascular diseases; /0393 model comprising a report system for screening tests)</p> <p>A61K49 (Preparations for testing in vivo: /0008 Screening agents using (transgenic) animal models or chimeric hosts, e.g. Alzheimer disease animal model, transgenic model for heart failure)</p>	<p>1572</p>	<p>217</p>	<p>4</p>

<p>Eggs and poultry dominated by A01K45/00.</p>	<p>A01K43 (/00 Testing, sorting or cleaning eggs, investigating or analysing eggs, e.g. by candling)</p> <p>A01K45 (/00 devices for determining whether a bird is about to lay; /005 harvesting or transport of poultry; /007 Injecting or otherwise treating hatching eggs)</p> <p>G01N33 (Analysing materials: /08 food > eggs, e.g. by candling; /085 by candling)</p>	<p>310</p>	<p>7</p>	<p>13</p>
<p>Peptidases dominated by C12Y304/21069 and C12N9/72.</p>	<p>C12N15 (Mutation or genetic engineering: /58 Recombinant DNA-technology > DNA or RNA fragments > Genes encoding for enzymes or proenzymes > Hydrolases > acting on peptide bonds > Plasminogen activators, e.g. urokinase, TPA)</p> <p>C12N9 (Enzymes, Proenzymes, Compositions thereof: /644 Proteinases, Endopeptidases > derived from animal tissue > from mammals > Serine endopeptidases > Coagulation factor IXa; /6456 plasminogen activators; /6559 t-plasminogen activator, i.e. tPA; /6462 u-Plasminogen activator, i.e. urokinase; /6464 Protein C (3.4.21.69); /72 peptidases > Urokinase)</p> <p>C12Y304 (Hydrolases acting on peptide bonds i.e. peptidases; /21022 Coagulation factor IXa; /21069 Protein C activated; /21073 u-Plasminogen activator (3.4.21.73), i.e. urokinase)</p>	<p>268</p>	<p>17</p>	<p>14</p>
<p>Vector systems, IGS from birds dominated by C12N2830/008 and C12N2800/30.</p>	<p>C07K16 (Immunoglobulins [IGs], e.g. monoclonal or polyclonal antibodies, antibodies with enzymatic activity: /02 from eggs)</p> <p>C07K2317 (Immunoglobulins specific features:</p>	<p>1133</p>	<p>186</p>	<p>6</p>

	<p>/23 from birds)</p> <p>C12N2800 (Nucleic acids vectors: /30 Vector systems comprising sequences for excision in presence of a recombinase, e.g. loxP or FRT)</p> <p>C12N2830 (Vector systems having a special element relevant for transcription: /008 cell type or tissue specific enhancer/promoter combination)</p>			
<p>Animals expressing milk dominated by A01K2267/01.</p>	<p>A01K2227 (Animals characterised by species: /10 mammal)</p> <p>A01K2267 (Animals characterised by purpose: /01 Animal expressing industrially exogenous proteins (Milk))</p> <p>A23C2230 (Aspects relating to animal feed or genotype: /05 milk products from transgenic animals)</p> <p>A23C9 (/00 Milk preparations)</p> <p>C07K14 (Peptides > 20 amino acids: /4732 from mammals > Casein in foodstuffs)</p> <p>C12N15 (Mutation or genetic engineering: /89 Introduction of foreign genetic material using processes not otherwise provided for, e.g. co-transformation > using microinjection)</p> <p>Y10S530 (Chemistry> peptides or proteins: /832 Proteins from mammals or birds > Milk, colostrum)</p>	1062	203	7
<p>Receptor antigens from animals dominated by C07K14/70596.</p>	<p>C07K14 (Peptides > 20 amino acids: /70539 from animals, from humans > Receptors, Cell surface antigens, Cell surface determinants, tumour specific antigens > Immunoglobulin superfamily ></p>	305	59	9

	<p>MHC-molecules, e.g. HLA-molecules; /70596 > Molecules with a "CD"-designation; /74 from animals, from humans > Major histocompatibility complex (MHC))</p> <p>G01N2333 (Assays involving biological materials from specific organisms or of a specific nature: /70596 from animals, from humans > Assays involving receptors, cell surface antigens or cell surface determinants > Molecules with a "CD"-designation)</p>			
T-Cell Receptors.	<p>C07K14 (Peptides > 20 amino acids: /7051 from animals, from humans > Receptors, Cell surface antigens, Cell surface determinants, tumour specific antigens > Immunoglobulin superfamily > T-cell receptor (TcR)-CD3 complex; /725 from animals, from humans > Receptors, Cell surface antigens, Cell surface determinants > T-cell receptors)</p>	97	10	12

Cleaning up after the Review

After reviewing the larger modularity classes of the remainder of the classification codes are searched with the aim of finding key classification classes as follows: A01K2207; A01K2217; A01K2227; A01K2267; A01K67; C07K14; C12N15; C12N2517; C12N5. The un-reviewed modularity classes typically consist on an orphan symbol. We inspect the definitions of the symbols in the modularity class and if necessary review the titles and abstracts for the corresponding patent publications. The results of this step are shown in Table 6.

Table 6 - Modularity classes identified by Clean Up

Summary	Key Classification Groups IPC and CPC	All Pubs	Latin Pubs	Graph Class
Genetically Modified Animals (Transgenic Pig)	A01K2217 (Genetically modified animals: /056 Animals modified by random mutagenesis > due to mutation of coding region of the transgene, dominant negative)	10	6	1742