Journal of Scientific and Technical Research
15 (1) 2025, 47-76
www.publishingindia.com/jstr

A Conceptual Framework of Cancer Prognosis Involving NGS Based DNA Fingerprinting

Afsheen Fatima[1], Farhan Ali[2] and Sher Ali[3*]
[1]Department of Biotechnology, Era University, Hardoi Road, Lucknow, Uttar Pradesh, India
Email: fatimaafsheen891@gmail.com
[2]Hospitalist, Internal Medicine, Yakima Valley, Memorial Hospital, WA, USA
Email: dr.farhanali30@gmail.com
[3]Vice Chancellor Office's, Era University, Hardoi Road, Lucknow, Uttar Pradesh, India
Email: profsali55@gmail.com
[*]Corresponding Author: profsali55@gmail.com

Abstract: The human body has close to about 200 different types of cells. Morphologically, and physiologically they are different. A perusal of literature shows that there are over 200 different types of cancers. These cancers are classified by the part of the body where they begin, and the type of cells they start in. It is therefore logical to infer that every single type of cell is prone to cancer. Once a cell becomes cancerous, it loses its Genetical, physical, physiological, immunological and morphological integrities. Each type of cancer has its own name and treatment. Common types of cancer are Bladder, Breast, Colorectal, Kidney, Lung, Lymphoma, Pancreatic, Prostate, Skin and uterine cancers. A major "burning question" in cancer research is "why do some people develop cancer while others with similar risk factors do not". This essentially boils down to understanding the complex interplay of genetics, environment, and lifestyle factors. These factors determine who gets cancer and who doesn't, even with similar exposures. About 5–10% of cancers are hereditary, meaning that they are caused by an inherited gene mutation. Thus, the risk of developing cancer is higher for people with inherited gene mutations. Several Cancers that have genetic basis include Breast, Bowel, Stomach and Prostate cancer besides Retinoblastoma in children. When the same types of cancers run in families, these are called familial cancer. In this paper, we discuss nuances of cancers and their correlation with different factors including lifestyle, food habits and environment. With this background, we present a Conceptual Framework of Cancer Prognosis Involving NGS Based DNA Fingerprinting. Such technical infusion is envisaged to uncover answers to a large number of questions facilitating not only prognosis but also deeper understanding of the aetiology of cancer and search for their likely remedial measures.

Keywords: Cancer, Cell types, Common cancer genes, DNA fingerprinting, Familial cancer, NGS sequencing.

I. Introduction

Seemingly, the topic dealing with Cancer Prognosis involving NGS Based DNA Fingerprinting seems to be less familiar. However, searching for a tool and reliable technique that can be used for an unequivocal diagnosis of cancer before it takes an ugly turn will be highly desirable. Cancer can occur in any part of the human body. For example, location wise, Breast cancer starts in the breast, Brain cancer starts in the brain and Lung cancer starts in the lung. Following this scheme, classification of cancers has been made (Table 1).

Table I: Types of Cancer Listed in the Literature

[Credit: https://www.cancer.gov/types]

Sr. No.	*Types of Cancer*	*Comments*
1	Acute Lymphoblastic Leukemia (ALL)	A fast growing cancer affecting blood and bone Marrow.
2	Acute Myeloid Leukemia (AML)	AML specifically targets myeloid cells.
3	Adolescents, Cancer	Includes lymphoma, leukemia, thyroid cancer and brain tumor.
4	Adrenocortical Carcinoma	A rare, aggressive cancer originates in the outer layer (cortex) of the adrenal glands and sits on top of the kidneys.
5	AIDS-Related Cancers	Have high risk of developing Kaposi sarcoma, non-Hodgkin lymphoma, and cervical cancer.

Sr. No.	Types of Cancer	Comments
6	Kaposi Sarcoma (Soft Tissue Sarcoma)	Kaposi sarcoma cancer cells are found in the skin or mucous membranes that line the gastrointestinal (GI) tract, from mouth to anus, including the stomach and intestines.
7	AIDS-Related Lymphoma (Lymphoma)	In case of AIDS-related lymphoma, malignant (cancer) cells form in the lymph system of patients who have acquired immunodeficiency syndrome (AIDS).
8	Primary CNS Lymphoma (Lymphoma)	It is a rare, aggressive form of non-Hodgkin lymphoma that starts in the brain, spinal cord, or meninges, and is characterized by malignant cells forming in the lymph tissue of the CNS.
9	Anal Cancer	This cancer forms in the tissues of the anus, often linked to human papillomavirus (HPV) infection and has the symptoms like bleeding, lumps, pain, or changes in bowel habits.
10	Appendix Cancer - see Gastrointestinal Neuroendocrine Tumors	This is rare and often discovered incidentally during appendectomy.
11	Astrocytomas, Childhood (Brain Cancer)	The most common type of brain tumor in children, originate from astrocytes.
12	Atypical Teratoid/Rhabdoid Tumor, Childhood, Central Nervous System (Brain Cancer)	Highly aggressive brain cancer, primarily affecting children under 3, with a poor prognosis.
13	Basal Cell Carcinoma of the Skin - see Skin Cancer	Basal Cell Carcinoma is a type of skin cancer; Atypical Teratoid/Rhabdoid Tumor (ATRT) is a rare, aggressive, and malignant brain cancer that primarily affects young children.
14	Bile Duct Cancer	While bile duct cancer (cholangiocarcinoma) is rare, brain metastases from it are exceptionally uncommon, and associated with a poor prognosis.
15	Bladder Cancer	This cancer starts in the lining of the bladder, often presents with blood in the urine (hematuria) and can be treated with surgery.
16	Bone Cancer (includes Ewing Sarcoma and Osteosarcoma and Malignant Fibrous Histiocytoma)	A rare but aggressive cancer that originates in bone cells.
17	Brain Tumors	Brain tumors are abnormal growths of cells within the brain or surrounding tissues.
18	Breast Cancer	Abnormal breast cells grow out of control and form tumours.
19	Bronchial Tumors (Lung Cancer)	Abnormal growths that can develop in the lining of the trachea.
20	Burkitt Lymphoma - see Non-Hodgkin Lymphoma	A rare, aggressive form of non-Hodgkin lymphoma, is a type of blood cancer affecting B-lymphocytes.
21	Carcinoma of Unknown Primary	(CUP) is a rare disease in which malignant (cancer) cells are found in the body but original place of the cancer began is not known.
22	Central Nervous System	Is the body's processing center, made up of the brain, spinal cord, and nerves.
23	Atypical Teratoid/Rhabdoid Tumor, Childhood (Brain Cancer)	This is a rare and aggressive brain cancer affecting children, often below 3 years and spread rapidly.
24	Medulloblastoma and Other CNS Embryonal Tumors, Childhood (Brain Cancer)	These tumors tend to spread through the cerebrospinal fluid to other parts of the brain and spinal cord.
25	Germ Cell Tumor, Childhood (Brain Cancer)	Develop from germ cells that fail to migrate to the ovaries or testes during fetal development, instead becoming trapped in the brain, often near the pineal or pituitary glands.

Sr. No.	*Types of Cancer*	*Comments*
26	Primary CNS Lymphoma	Uncommon but fast growing lymphoma.
27	Cervical Cancer	Growth of cells that starts in the cervix, can be prevented by vaccine.
28	Childhood Cancers	A group of cancers that occur in children from birth to age 14, and are distinct from adult cancers in their type and growth.
29	Childhood Cardiac Tumors Treatment	Watchful waiting, surgery, chemotherapy, radiation therapy, or a combination of these approaches.
30	Cancers of Childhood, Rare	All childhood cancers are rare, some are particularly infrequent, like pancreatoblastoma, malignant Rhabdoid and Pleuropulmonary Blastoma.
31	Cholangiocarcinoma - see Bile Duct Cancer	A rare type of cancer that develops in the bile ducts, the small tubes that carry bile from the liver and gallbladder to the small intestine.
32	Chordoma, Childhood (Bone Cancer)	A rare bone cancer that arises from remnants of the notochord, a structure that forms the spine during development.
33	Chronic Lymphocytic Leukemia (CLL)	CLL is a slow-growing cancer affecting blood and bone marrow, specifically targeting B lymphocytes.
34	Chronic Myelogenous Leukemia (CML)	CML is the cancer of e bone marrow produces too many abnormal WBC specifically granulocytes.
35	Colorectal Cancer	Cancer that develops in the tissues of the colon or rectum.
36	Craniopharyngioma, Childhood (Brain Cancer)	Is a rare, benign (non-cancerous) brain tumor that typically arises near the pituitary gland in children.
37	Cutaneous T-Cell Lymphoma - see Lymphoma (Mycosis Fungoids and Sézary Syndrome)	(CTCL) is a type of non-Hodgkin lymphoma that affects the skin.
38	Ductal Carcinoma In Situ (DCIS) - see Breast Cancer	A non-invasive, pre-invasive breast cancer where abnormal cells are confined to the milk ducts and haven't spread to surrounding breast tissue.
39	Diffuse Intrinsic Pontine Glioma (DIPG) (Brain Cancer)	A highly aggressive, malignant brain tumor that occurs in the brainstem's pons often resulting death.
40	Embryonal Tumors, Medulloblastoma and Other Central Nervous System, Childhood (Brain Cancer)	May begin in embryonic (fetal) cells that remain in the brain after birth.
41	Endometrial Cancer (Uterine Cancer)	Often curable, especially when detected early, with surgery (hysterectomy).
42	Ependymoma, Childhood (Brain Cancer)	This cancer arises from ependymal cells lining the ventricles and central canal of the brain and spinal cord, more common in children.
43	Esophageal Cancer	A malignancy that originates in the esophagus, the tube connecting the throat to the stomach.
44	Esthesioneuroblastoma (Head and Neck Cancer)	Esthesioneuroblastoma, also called olfactory Neuroblastoma. This malignant tumor originates in the upper part of the nasal cavity, specifically the olfactory epithelium, which is responsible for our sense of smell.
45	Ewing Sarcoma (Bone Cancer)	A rare, aggressive bone cancer affecting arms and legs.
46	Extracranial Germ Cell Tumor, Childhood	This originates from germ cells in areas of the body other than the brain, such as the testicles, ovaries, mediastinum, retro peritoneum, sacrum, coccyx, head and neck.
47	Extragonadal Germ Cell Tumor	Rare cancers that develop from germ cells, (sperm and eggs), but remain outside the gonads.
48	Eye Cancer	Ocular cancer, encompasses intraocular melanoma, intraocular lymphoma, and retinoblastoma.
49	Intraocular Melanoma	Serious cancer that develops in the cells that produce pigment within the eye, (such as iris, Ciliary body, or choroid).

Sr. No.	*Types of Cancer*	*Comments*
50	Retinoblastoma	A rare, malignant tumor of the retina that occurs in children.
51	Fallopian Tube Cancer	A rare form of gynecologic cancer.
52	Gallbladder Cancer	develops when abnormal cells in the gallbladder grow out of control and form a tumor.
53	Gastric (Stomach) Cancer	Is a malignant tumor that develops in the stomach lining, with the most common type being adenocarcinoma.
54	Gastrointestinal Neuroendocrine Tumors	Also known as carcinoid tumors, originates from neuroendocrine cells in the digestive system.
55	Gastrointestinal Stromal Tumors (GIST) (Soft Tissue Sarcoma)	Rare cancers that develop in the digestive system, specifically in the walls of the gastrointestinal tract.
56	Germ Cell Tumors	Arising from germ cells that can be benign or malignant, most commonly occurring in the testicles or ovaries, but also in other areas like the brain, abdomen, and chest.
57	(Soft Tissue Sarcoma)	A rare cancer that forms in the soft tissues of the body
58	Childhood Central Nervous System Germ Cell Tumors (Brain Cancer)	Childhood CNS germ cell tumors include unusual thirst, frequent urination, or vision changes.
59	Childhood Extracranial Germ Cell Tumors	This originates from developing sperm or egg cells that travel to parts of the body other than the brain.
60	Ovarian Germ Cell Tumors	Rare tumors that arise from the reproductive cells (germ cells) in the ovaries.
61	Testicular Cancer	A cancer that develops in the testicles.
62	Gestational Trophoblastic Disease	A group of rare pregnancy-related tumors arising from abnormal proliferation of trophoblastic tissue.
63	Hairy Cell Leukemia	A rare, blood cancer characterized by abnormal B lymphocytes affecting the bone marrow, spleen, and blood.
64	Head and Neck Cancer	A group of cancers of the mouth, sinuses, nose or throat.
65	Heart Tumors, Childhood	Rare in children, and most are benign (non-cancerous).
66	Hepatocellular (Liver) Cancer	Primary liver cancer, developing from hepatocytes and often occurring in people with chronic liver diseases like cirrhosis or hepatitis B or C infections.
67	Histiocytosis, Langerhans Cell	A rare disorder where body produces too many immature Langerhans cells, damage tissues and organs, such as skin and bones.
68	Hodgkin Lymphoma	A rare cancer that affects the lymphatic system, which is part of the immune system.
69	Hypopharyngeal Cancer (Head and Neck Cancer)	Most common type, develops in the hypopharynx, the lower part of the throat, and is often associated with smoking and alcohol use.
70	Intraocular Melanoma	A rare but serious one, develops in the pigment producing cells within the eye.
71	Islet Cell Tumors, Pancreatic Neuroendocrine Tumors	Rare cancers that arise from the hormone-producing cells (islet cells) of the pancreas.
72	Kidney (Renal Cell) Cancer	Common type of kidney cancer, originating in the lining of the kidney's tiny tubes (renal tubules).
73	Leukemia	A cancer of the blood and bone marrow, which can impair the production of healthy blood cells.
74	Lip and Oral Cavity Cancer (Head and Neck Cancer)	Typically starts in the thin, flat cells lining the lips and mouth, knownas squamous cell carcinomas, and can spread into the deeper tissues.
75	Liver Cancer	A serious condition arises from long-term liver damage often linked with cirrhosis and hepatitis B or C infection.

Sr. No.	*Types of Cancer*	*Comments*
76	Lung Cancer (Non-Small Cell, Small Cell, Pleuropulmonary Blastoma, Pulmonary Inflammatory Myofibroblastic Tumor, and Tracheobronchial Tumor)	This includes adenocarcinoma, squamous cell carcinoma, large and small cell carcinoma, and pleuropulmonary blastoma, pulmonary inflammatory myofibroblastic and tracheobronchial tumors.
77	Lymphoma	A type of blood cancer that affects the immune system.
78	Male Breast Cancer	This originates in the breast tissue of men, accounting for less than 1% of all breast cancer cases.
79	Melanoma	Melanoma is a skin disease in which malignant cells form in melanocytes.
80	Melanoma, Intraocular (Eye)	A rare but serious cancer that develops in the cells that produce pigment within the eye.
81	Merkel Cell Carcinoma (Skin Cancer)	Rare but aggressive skin cancer that originates in Merkel cells, a type of neuroendocrine cell in the skin's epidermis, often appearing as a rapidly growing, firm bump on sun-exposed skin.
82	Mesothelioma, Malignant	A rare but aggressive cancer that develops in the mesothelium, strongly linked to asbestos exposure.
83	Metastatic Cancer	Cancer that spreads from its original site to other parts of the body.
84	Metastatic Squamous Neck Cancer with Occult Primary (Head and Neck Cancer)	Squamous cell cancer that has spread to lymph nodes in the neck.
85	Midline Tract Carcinoma With NUT Gene Changes	A rare, aggressive cancer characterized by a chromosomal rearrangement involving the NUT gene, often leading to fusion proteins that disrupt normal cell function.
86	Mouth Cancer (Head and Neck Cancer)	Encompass a range of cancers that originate in the tissues of the head and neck, including the mouth, throat, nose, and salivary glands.
87	Multiple Endocrine Neoplasia Syndromes	Inherited disorders that affect the endocrine system.
88	Multiple Myeloma/Plasma Cell Neoplasms	Inherited disorders that affect the endocrine system.
89	Mycosis Fungoides (Lymphoma)	The most common type of cutaneous T-cell lymphoma (CTCL), a type of blood cancer that affects the skin.
90	Myelodysplastic Syndromes, Myelodysplastic/Myeloproliferative Neoplasms	A group of blood cancers that exhibit features of both myelodysplastic syndromes (MDS) and myeloproliferative neoplasms (MPN), characterized by abnormal blood cell production and/or overproduction.
91	Myelogenous Leukemia, Chronic (CML)	A blood cancer where bone marrow produces too many white blood cells, specifically granulocytes, which can lead to various health problems.
92	Myeloproliferative Neoplasms	A group of diseases in which the bone marrow makes too many red blood cells, white blood cells, or platelets.
93	Nasal Cavity and Paranasal Sinus Cancer (Head and Neck Cancer)	Originates in the nasal cavity and the air-filled spaces around the nose, with squamous cell carcinoma being the most common type.
94	Nasopharyngeal Cancer (Head and Neck Cancer)	Originates in the nasopharynx, the upper part of the throat behind the nose, and is often linked to Epstein-Barr virus (EBV) infection.
95	Neuroblastoma	A rare cancer that develops in nerve tissue.
96	Neuroendocrine Tumors (Gastrointestinal)	A type of neuroendocrine tumor that develops within the GI tract, anywhere from the esophagus to the anus.
97	Non-Hodgkin Lymphoma	A cancer that develops in the lymphatic system, affecting lymphocytes.

Sr. No.	*Types of Cancer*	*Comments*
98	Non-Small Cell Lung Cancer	NSCLC, accounting for about 85% of all lung cancers.
99	Neuroendocrine Tumors (Gastrointestinal)	Rare one originates from neuroendocrine cells in the digestive system, often in the small intestine, rectum, or appendix.
100	Osteosarcoma and Undifferentiated Pleomorphic Sarcoma of Bone Treatment	Treatment for both osteosarcoma and undifferentiated pleomorphic sarcoma (UPS) of bone involves surgery, chemotherapy, and radiation therapy.
101	Ovarian Cancer	Ovaries grow and multiply uncontrollably, forming a tumor, and can spread to other parts of the body.
102	Pancreatic Cancer	Pancreatic cancer, forms in the tissues of the pancreas, difficult to diagnose.
103	Pancreatic Neuroendocrine Tumors (Islet Cell Tumors)	These are rare tumors that originate in the hormone-producing cells of the pancreas, and can be either benign or cancerous.
104	Papillomatosis (Childhood Laryngeal)	A benign, viral-induced condition giving rise to wart-like growths in the respiratory tract, most commonly the larynx (voice box), causing airway obstruction.
105	Paraganglioma	Paraganglioma is a rare tumor that forms near nerves and blood vessels outside of the adrenal glands.
106	Paranasal Sinus and Nasal Cavity Cancer (Head and Neck Cancer)	This malignant (cancer) cells form in the tissues of the paranasal sinuses and nasal cavity.
107	Parathyroid Cancer	A rare cancer that occurs in the parathyroid glands in the neck.
108	Penile Cancer	This malignancy affects the skin of the penis, with squamous cell carcinoma being the most common type.
109	Pheochromocytoma	A type of neuroendocrine tumor that grows from cells called chromaffin cells.
110	Pituitary Tumor	An abnormal growth of cells within the pituitary gland.
111	Plasma Cell Neoplasm/Multiple Myeloma	are cancers that arise from abnormal plasma cells (a type of white blood cell) that build up in the bone marrow and form tumors, often leading to bone damage and other complications.
112	Pleuropulmonary Blastoma (Lung Cancer)	A rare and aggressive type of childhood cancer that forms in the lungs, the tissue that covers the lungs, and the inside wall of the chest cavity called the pleura.
113	Pregnancy and Breast Cancer	Breast cancer during pregnancy is rare, occurring in about 1 in 3,000 pregnancies.
114	Pregnancy and Hodgkin Lymphoma	Hodgkin lymphoma (HL), a type of cancer, can occur during pregnancy, but it's rare, with around 1 in 6,000 pregnancies.
115	Pregnancy and Non-Hodgkin Lymphoma	When NHL is diagnosed during pregnancy, it is fast-growing (high-grade) type.
116	Primary Central Nervous System (CNS) Lymphoma	Is a rare, aggressive form of non-Hodgkin lymphoma that originates in the brain, spinal cord, leptomeninges, or eyes.
117	Primary Peritoneal Cancer	a rare cancer that originates in the peritoneum, the tissue lining the abdominal cavity.
118	Prostate Cancer	Prostate cancer is a disease where uncontrolled cells grow in the prostate gland.
119	Pulmonary Inflammatory Myofibroblastic Tumor (Lung Cancer)	IMT of the lung represents an extremely rare type of tumor that appears most commonly in children and young individuals.
120	Rare Cancers of Childhood	Pancreatoblastoma, malignant rhabdoid tumors, and melanotic neuroectodermal tumors of infancy.
121	Rectal Cancer	Rectal cancer develops in the rectum, can be treated through surgery, radiation, and chemotherapy.

Sr. No.	Types of Cancer	Comments
122	Recurrent Cancer	Occurs when some cancer cells survive initial treatment and grow again, either in the original location or elsewhere in the body.
123	Renal Cell (Kidney) Cancer	Renal cell cancer is a type of cancer that forms in tubules of the kidney.
124	Retinoblastoma	A rare, malignant tumor of the retina that occurs in children.
125	Rhabdomyosarcoma, Childhood (Soft Tissue Sarcoma)	A cancer that develops from primitive muscle cells, and can occur anywhere in the body, but most commonly around the head and neck, bladder, or testes in boys.
126	Salivary Gland Cancer (Head and Neck Cancer)	Originates in the salivary glands, which produce saliva, and can manifest as a lump or swelling in the face or neck, or difficulty swallowing or opening the mouth.
127	Sarcoma	Sarcoma is a rare type of cancer that develops in the bones and soft tissues.
128	Childhood Rhabdomyosarcoma (Soft Tissue Sarcoma)	A malignant tumor of mesenchymal origin that can occur in various organs and tissues, including those without striated muscle.
129	Ewing Sarcoma (Bone Cancer)	A rare, aggressive type of bone cancer that primarily affects children and young adults.
130	Osteosarcoma (Bone Cancer)	Common type of primary bone cancer, originating in cells that form new bone tissue, affects teenagers and young adults.
131	Childhood Rhabdomyosarcoma (Soft Tissue Sarcoma)	This malignant tumor originating from primitive mesenchymal cells that typically differentiate into skeletal muscle tissue.
132	Childhood Vascular Tumors (Soft Tissue Sarcoma)	Abnormal growths of blood vessels or lymph vessels that can be either benign or malignant.
133	Ewing Sarcoma (Bone Cancer)	A rare, aggressive type of bone cancer that primarily affects children and young adults.
134	Uterine Sarcoma	Uterine sarcoma is a rare cancer that originates in the muscle or supporting tissues of the uterus.
135	Skin Cancer	Skin cancer, a disease where abnormal cells grow in the skin.
136	Small Cell Lung Cancer	SCLC is an aggressive, fast-growing type of lung cancer that often starts in the central airways and tends to spread rapidly.
137	Small Intestine Cancer	A rare malignancy that develops in the small intestine.
138	Squamous Neck Cancer with Occult Primary, Metastatic (Head and Neck Cancer)	Squamous cell cancer spreads to lymph nodes in the neck, but the original (primary) cancer site within the body is unknown.
139	Stomach (Gastric) Cancer	Develops when cells in the stomach grow out of control, with most cases being adenocarcinomas originating in the stomach's inner lining.
140	T-Cell Lymphoma, Cutaneous - see Lymphoma (Mycosis Fungoides and Sèzary Syndrome)	These types of non-Hodgkin lymphoma primarily affect the skin.
141	Nasopharyngeal Cancer	This is a type of head and neck cancer that develops in the nasopharynx, the upper part of the throat behind the nose.
142	Throat Cancer (Head and Neck Cancer)	Head and neck cancer includes epithelial malignancies of the upper aerodigestive tract (UADT), including the paranasal sinuses, nasal cavity, oral cavity, pharynx, and larynx.
143	Oropharyngeal Cancer	Is a type of head and neck cancer that develops in the tissues of the oropharynx.
144	Hypopharyngeal Cancer	Rare type of throat cancer, develops in the lower part of the pharynx, the area where the larynx and esophagus meet, and is often a squamous cell carcinoma.

Sr. No.	*Types of Cancer*	*Comments*
145	Thymoma and Thymic Carcinoma	Rare cancers that arise from the epithelial cells of the thymus, a gland in the upper chest.
146	Thyroid Cancer	Malignant cells form in the thyroid gland.
147	Tracheobronchial Tumors (Lung Cancer)	Rare one and can cause breathing problems by obstructing the airway.
148	Transitional Cell Cancer of the Renal Pelvis and Ureter (Kidney (Renal Cell) Cancer)	A type of cancer that originates in the transitional cells lining the urinary tract, including the renal pelvis and ureters.
149	Urethral Cancer	A rare malignancy, develops in the urethra, the tube that carries urine out of the body.
150	Uterine Cancer, Endometrial	A type of cancer that starts in the lining of the uterus (endometrium).
151	Vaginal Cancer	very rare malignancy, accounting for 1-2% of all female genital tract cancers, and is often linked to HPV infection.
152	Vulvar Cancer	A rare cancer affecting the vulva (external female genitalia), often presents as a lump, sore, or skin changes like itching, burning, or bleeding.
153	Vascular Tumors (Soft Tissue Sarcoma)	Vascular sarcoma originates from the blood vessels or lymphatic vessels and can affect various body parts.
154	Wilms Tumor and Other Childhood Kidney Tumors	Wilms tumor, also known as nephroblastoma, is the most common type of kidney cancer in children.

Table I shows different types of cancers, their origin and severity. However, with respect to actual numbers and types of cancers, there is no guarantee that this table is complete. The purpose of this table is to highlight the facts that some cancers are very similar with respect to their etiology and therefore accurate diagnosis may pose a challenge. In view of this, their prognosis using the NGS DNAF approach needs to be carefully calibrated.

In this paper, we briefly summarize the genesis, concept and execution of DNA fingerprinting (Jeffreys *et al.,* 1985; Ali *et al.,* 1986; Afsheen *et al.,* 2025), its gradual advancement and eventual automation. Using desired cancer biopsied samples or established cell lines, we will isolate genomic and cDNA from these sources. Following this, we will focus on allelic alteration of genes using NGS sequencing employing DNA fingerprinting approach. The only difference here will be that the DNA and cDNA will be used from cancer cells. For control, DNA from normal healthy individuals will be used. For fingerprinting, well characterized STRs would be used (Norrgard, 2008).

Finally, based on the NGS sequencing, (See Next-Generation Sequencing: Advantages, Disadvantages, and Future. Available from: https://www.researchgate.net/publication/306431728_NextGeneration_Sequencing_Advantages_Disadvantages_and_Future [Accessed Sep 05 2018]) we would highlight how genes, intronic sequences or repetitive DNA show allelic alteration amongst cancer samples compared to that in normal ones. Allelic variation has been extensively applied for DNA fingerprinting by law enforcement agencies to nail the culprits. It may be noted that these alleles may be the representatives of a functional genes or related to variable number tandem repeat (VNTR) loci, short tandem repeat (STR) also referred to as minisatellites located close to a functional gene (Pathak *et al.,* 2006; Srivastava *et al.,* 2008; Srivastava *et al.,* 2009). For scoring allelic variation, if the cDNA is used as a template, they may pick up alleles corresponding to the functional genes (Pathak and Ali, 2011). There is yet another technical approach to undertake molecular mining of altered genes from the cancer cells. This approach is known as minisatellite associated sequence amplification (MASA) (Bashamboo and Ali, 2001). It works both with genomic DNA and cDNA giving rise to a large number of bands after the single primer reaction is resolved on the agarose gel (Ali *et al.,* 1999). DNAF or MASA approaches have not been exploited for cancer studies although both genomic and cDNA from normal genome shave been used successfully for MASA reaction (Ali *et al.,* 1999; Bashamboo *et al.,* 1999). The technical difference between MASA and STR based allelic variation is that in case of MASA reaction, only a single primer is used whereas in STR a set of primers is used. As an example, we have shown here THO1 locus its chromosomal location and 6 sets of primers with their sequences and PCR product size.

Table II: Details of TH01 Locus with 6 Sets of Primers

Other Names	*Chromosomal Location*	*GenBank Accession*
HUMTH01, TC11 UniSTS: 240639	11p15.5; intron 1 of human tyrosine hydroxylase gene Chr 11; 2.149 Mb (May 2004, NCBI build 35)	D00269; has 9 repeat units

Repeat: [AATG] = bottom strand (commonly used); [TCAT] = GenBank top strand

Reported Primers	*Ref.*	*PCR Primer Sequences*
Set 1	1, 3	5'-GTGGGCTGAAAAGCTCCCGATTAT-3' (AATG strand) 5'-ATTCAAAGGGTATCTGGGCTCTGG-3' (TCAT strand) PCR Product size:171 bp both for 1 and 3
Set 2	4	5'-GTGGGCTGAAAAGCTCCCGATTAT-3' (AATG strand) 5'-GTGATTCCCATTGGCCTGTTCCTC-3' (TCAT strand) PCR Product size: 146 bp
Set 3	125	5'-GCTTCCGAGTGCAGGTCACA-3' (AATG strand) 5'-CAGCTGCCCTAGTCAGCAC-3' (TCAT strand) PCR Product size: 230 bp
Set 4	Promega	PowerPlex 1.1 (TMR labeled) primer sequences 5'-ATTCAAAGGGTATCTGGGCTCTGG-3' 5'-[TMR]-GTGGGCTGAAAAGCTCCCGATTAT-3' PCR Product size: 171
Set 5	ABI	COfiler (JOE labeled), SGM Plus (NED labeled), Identifiler (VIC labeled) PCR Product size : 160
Set 6	Promega	PowerPlex 2.1, PowerPlex 16 (FL labeled) primer sequences 5'-[FL]-GTGATTCCCATTGGCCTGTTC-3' 5'-ATTCCTGTGGGCTGAAAAGCTC-3' PCR Product size: 152

For details [See http://www.nist.gov/public_affairs/privacy.cfm.].

We propose to focus on allelic variation using cancer genomic and cDNA employing direct NGS sequencing using well defined STRs. Cancer DNA and cDNA analysis employing the MASA approach will be part of another publication. Our proposed conceptual framework of Cancer Prognosis using NGS Based DNA Fingerprinting will involve cancer biopsied samples for DNA and mRNA isolation, STRs for reaction, NGS machine for sequencing coupled with software and cDNA information of all the genes that have been unequivocally implicated with cancer etiology. We call it a conceptual framework because this approach has yet to be used for studying allelic variation of sequences due to cancer. In order to uncover such altered genes or otherwise, we propose to use available DNA markers based on STRs totaling up to 68 that are available globally for forensic DNA fingerprinting. From amongst these STR markers, 20 are routinely used world over which have become the basis of CODIS relied upon heavily by law enforcement agencies. In extended tests, these 68 DNA STR markers, compared to the industry standard of only 16 DNA STR markers have shown a 99.9999% accuracy of a result. Testing for fewer markers may appear to be more economical but the results may become inconclusive. In the following paragraph, we have briefly discussed how the human genome is organized (Kumari *et al.,* 2015) and satellite tagged genes are expressed (Pathak *et al.,* 2010; Pathak and Ali, 2011).

The human genome contains approximately 20000-22000 genes with ~3.5 billion haploid sequences. The actual number of genes present in the human genome will never ever be known with the available technology. This is because the human genome is highly heterogeneous and no two person's genomes or expressed genes are identical (Ali and Gangadharan, 2000). Another reason is that a large number of genes undergo alternative splicing giving rise to several mRNA transcripts (Pathak *et al.,* 2006). These transcripts are independently translated giving rise to multiple proteins. Thus, one gene one protein concept is not operative. Yet another mechanism making the human genome highly dynamic is the gene conversion and its copy number variation. Owing to the presence of repeat sequences that tend to shrink and expand, gene death and birth remain a continuous process (Pathak *et al.*, 2006). This is yet

another cause of genomic heterogeneity. If a given Genetic alteration is favoured by the forces of nature, the newly born gene is fixed in the genome (allele fixation). However, if the altered gene is not favoured by the forces of evolution, gene death is ensured following the Darwinian law of the survival of the fittest (Darwin, 1859). How our understanding about genome organization and gene expression would help us to decipher CODIS based allelic variation employing NGS is highlighted in this paper. Similarly, DNA markers and CODIS are also discussed.

II. DNA Markers

DNA markers, also known as genetic or molecular markers are specific DNA sequences with known chromosomal locations used to identify genes, detect genetic differences, and study diseases (Bashamboo *et al.,* 2001). They serve as "flags" or "landmarks" that help researchers pinpoint genes or regions of genes of interest. There are various types of DNA markers, including: Single Nucleotide Polymorphisms (SNPs) showing Variations at a single DNA base (A, C, G, or T). The next is Microsatellites or Short, repetitive DNA sequences (Ali *et al.,* 1986; Schaefer *et al.,* 1986; John *et al.,* 1996). In addition, we use Restriction Fragment Length Polymorphisms (RFLPs) as markers. RFLP shows differences in DNA fragment lengths caused by variations in restriction enzyme recognition sites (Yam *et al.,* 1987; Ali and Wallace, 1988). Similarly, we find Amplified Fragment Length Polymorphisms (AFLPs). This shows differences in DNA fragment lengths after PCR amplification of restriction fragments. In addition, we use Random Amplified Polymorphic DNA (RAPDs) which shows DNA fragments amplified using random primers. In RAPDs, the primers are random therefore the result will be inconsistent owing to several technical reasons. For example, if the primer is replaced with another sequence complexity, the resultant band profile will also be changed.

III. Applications

In Genetical studies, marker systems are employed for a variety of purposes. These are discussed hereunder.

Gene Mapping: Identifying the location of genes on chromosomes.

Genetic Variation Detection: Identifying differences in DNA sequences between individuals or populations.

Disease Studies: Studying the genetic basis of diseases and identifying genes associated with specific conditions.

Species Identification: Distinguishing between different species or varieties.

Plant Breeding: Identifying desirable traits in plants and selecting for them.

Forensic Science: Identifying individuals based on DNA profiles.

Personalized Medicine: Identification of individuals with specific genetic predispositions to a disease.

IV. Ideal Characteristics of DNA Markers

Polymorphism: The marker should have multiple alleles (different versions of the marker) to allow for the identification of genetic variations. The polymorphism is related to sequence alteration (mutations). This mutation may occur in coding or non-coding regions of the gene. It may be noted that occurrences of mutation resulting in polymorphism and their scoring (detection) are not always 100%. This means some allelic variation may be missed out. This is true more in experiments where synthetic Oligodeoxyribonucleotides are used. A single point mutation in the target DNA will not allow synthetic DNA to bind with target DNA in a PCR resulting in reaction failure. Thus, it is agreed upon that several markers should be used for DNA fingerprinting and conclusion should be based on collective results. In view of this technical nuance, CODIS systems have been developed that are useful for forensic application.

CODIS, which stands for the *Combined DNA Index System,* is a United States national DNA database maintained by the FBI. This is used to link crimes and identify potential suspects by matching DNA profiles from crime scenes with those of convicted offenders. This began as a pilot project in 1990 overseeing 14 states of the USA and local laboratories. The DNA identification act of 1994 (public Law 103-322) formalized the FBI's authority to establish a national DNA index marker for law enforcement purposes. Eventually, the FBI's National DNA Index System (NDIS) became operational.

Abundance and Uniform Distribution: The marker should be present in multiple locations throughout the genome to facilitate the identification of genes of interest.

Co-Dominance: The marker should allow for the identification of heterozygotes (individuals with two different alleles of the marker).

Easy to Genotype: The marker should be easy and cost-effective to analyse. Full articles may be found in DNA molecular markers in plant breeding; current status and recent advancements in genomic selection and genome editing.

The places where DNA commonly varies between people are called single nucleotide polymorphisms or "SNPs" for short (pronounced "snips"). Ancestry DNA scientists have carefully chosen more than 700,000 DNA markers across our DNA to analyse.

There are ten common nDNA, mtDNA, cpDNA, pDNA, eDNA, cfDNA, ctDNA, rDNA, rcDNA, and cccDNA terminologies that are used to explain various types of DNA. Here. Each type of DNA, with their full form and function are listed.

Table III: Full Form of 10 Different Types of DNA

Sr. No.	*Types of DNA*	*Description*
1	nDNA	The full form of nDNA is nuclear DNA or nuclear deoxyribose nucleic acid.
2	mtDNA	The full form of mtDNA is mitochondrial DNA or mitochondrial deoxy-ribonucleic acid.
3	cpDNA	The full name of cpDNA is chloroplast DNA or chloroplast deoxyribose nucleic acid.
4	pDNA	The full name of pDNA is plasmid DNA.
5	eDNA	eDNA: The full name of eDNA is environmental DNA.
6	cfDNA	The full name of cfDNA is cell-free DNA or cell-free fetal DNA.
7	ctDNA	The full name of ctDNA is circulating tumor DNA.
8	rDNA	The full name of rDNA is recombinant DNA.
9	rcDNA	The full name of rcDNA is relaxed circular DNA.
10	cccDNA	The full name of cccDNA is covalently closed circular DNA.

[Credit: https://geneticeducation.co.in/full-form-of-10-types-of-dna/]

The Combined DNA Index System (CODIS) uses a core set of 20 Short Tandem Repeat (STR) markers for forensic DNA analysis, comprising 13 originally listed and 7 added in 2017. The other DNA markers available reach up to 64 for CODIS besides (SNP) single nucleotide polymorphism (https://oig.justice.gov/reports/FBI/a0632/final.pdf). However, we have listed only 20 markers that are commonly used for forensic application.

Table IV: Details of 20 STRs and Their Chromosomal Location

Sr. No.	*Original 13 CODIS Core Loci*	*Chromosomal Location*
1	CSF1PO	CSF1PO locus resides on chromosome 5
2	D3S1358	Short arm (p) of chromosome 3, at position 3p21.
3	D5S818	STR D5S818 is at 5q23.2 region
4	D7S820	STR D7S820 is at 7q21.11 region
5	D8S1179	STR D8S1179 is at the 8q24.13 region
6	D13S317	STR D13S317 is at 13q22-q31 region
7	D16S539	STR D16S539 is at 16q24.1 region
8	D18S51	STR D18S51 is at 18q21.3 region
9	D21S11	STR D21S11 is at 21q21.1. region
10	FGA	STR FGA is at 4q31.3.
11	TH01	STR TH01 is at 11p15.5, region
12	TPOX	STR TPOX is at 2p25.3. region
13	vWA	STR vWA is at 12p12-12pter region.
Additional 7 CODIS Core Loci (Added in 2017)		
14	D1S1656	STR marker D1S1656 is located on chromosome 1
15	D2S441	STR marker D2S441 is at 2p14 region
16	D2S1338	STR maker D2S1338 is at 2q35 region
17	D10S1248	STR marker D10S1248 is at 2q35 region
18	D12S391	STR marker D19S433 is at 12p 341p region
19	D19S433	STR marker D19S433 is located on chromosome 19
20	D22S1045	STR D22S1045 is located at 22q12.3 region

STRs are named for the gene locus with which they are physically associated (e.g., VWA, is associated with 40th intron of the von Willebrand Factor gene) and located on chromosome 12p. Similarly, STR locus*D13S317is located on chromosome 13. Here "D" for DNA, "13" for chromosome 13, "S" for STR, and "317" is a unique identifier). However, sometimes, it is not apparent. For example, CSF1PO location is not clear from the table. Literature analysis shows that this allele (CSF1PO) is located on chromosome 5. Every single allele which is part of CODIS has a history and has been included in the list of forensic markers after a detailed test and quality control. Complete locus information for CSF1PO is given hereunder as an example.

Table V: Locus Information of _CSF1PO

Chromosomal Location	*5q33.3-34*
Genebank Locus and Locus Definition	HUMCSF1PO Human c-fms proto-oncogene for CSF-1 receptor gene
Repeat Sequence 5′ – 3′	AGAT(*)
Fluorecent Label	*TMR:* 5'-terminal carboxy-tetramethylrhodamine label
Size Range of allelic ladder components(**, ****) (bases)	291-327
Repeat Numbers of allelic ladder components	6-15(**)
Repeat Numbers of alleles not present in allelic ladder	None
Commercial Product	PowerPlexTM 1.2 System of the Promega Co.

V. Another Example of STR Locus "D1S1656" and Its Interpretation

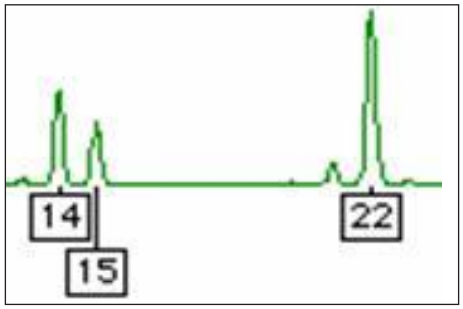
Locus "D1S1656" refers to a specific short tandem repeat (STR) on chromosome 1 commonly used in forensic DNA analysis and paternity testing.

Here's a more detailed breakdown

- *"D":* Indicates a DNA locus.
- *"1":* Refers to chromosome 1.
- *"S":* Stands for single copy STR sequence.
- *"1656":* Is the unique identifier for this specific locus.
- *STR Locus:* D1S1656 is a locus characterized by short tandem repeats (STRs), which are markers of choice for human forensic identification.
- *Forensic Applications:* D1S1656 is included in widely-used multiplex kits from several vendors and is used for human forensic identification because of its high polymorphism, relatively low mutation rate, and amplification efficiency.
- *Paternity Testing:* D1S1656 is also used in paternity testing.
- *Allele Distribution:* Allele distribution and genetic parameters for D1S1656 have been studied in various populations.

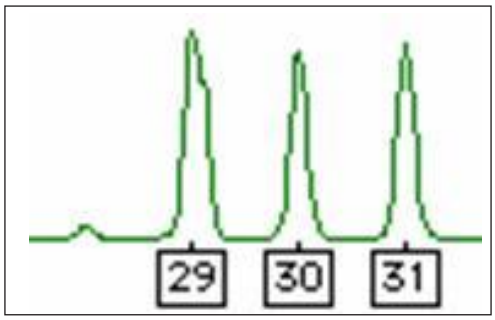
VI. Now We Describe Tri-Allelic Patterns

Tri-Allelic is important because three-banded or tri-allelic patterns are sometimes observed at a single locus in a multiplex STR profile. The table of tri-allelic patterns is useful to "compare notes" and see the allelic profile and their similarity. Tri-allelic patterns generally fall into one of the two different groups based on relative peak heights as shown hereunder.



Type 1

Type 1: Tri-allelic pattern here showing that sum of the heights of two peaks is equal to the third one. If there is a gene duplication at a given locus which is specific to a given STR, that will also show three bands. If the gene duplication has taken place at the same locus more than once, the corresponding number of bands will be seen. These peaks were obtained using NGS sequencing system.



Type 2: Balanced Peak Heights

Table VI: Examples of STR Useful for Tri-allelic Pattern

Sr. No.	*Core STR Loci (272)*	*Other Common STR Loci (65)*	*Y-STR Loci (64) Duplication or Triplication*
1	CSF1PO (9)	D2S1338 (10)	DYF387S1 (1)
2	FGA (40)	D19S433 (12)	DYS19 (8)
3	TH01 (5)	Penta D (12)	DYS389I (4)
4	TPOX (19)	Penta E (16)	DYS389II (6)
5	VWA (26)	F13A01	DYS390 (1)
6	D3S1358 (11)	FES/FPS (1)	DYS391 (3)
7	D5S818 (8)	F13B	DYS392
8	D7S820 (22)	LPL	DYS393 (2)
9	D8S1179 (22)	SE33 (6)	DYS385 a/b (19)
10	D13S317 (16)	D10S1248 (2)	DYS438 (4)
11	D16S539 (13)	D12S391 (2)	DYS439 (5)
12	D18S51 (44)	D22S1045 (1)	DYS437 (2)
13	D21S11 (27)	D1S1656 (3)	DYS448 (3)
14		D2S441	DYS456 (2)
15			DYS458 (3)
16			DYS533 (1)
17			DYS570 (1)
18			DYS576 (3)
19			DYS635 (1)
20			Y-GATA-H4 (1)

Credit: https://strbase-archive.nist.gov/tri_tab.htm

VII. Combined DNA Index System (CODIS)

Combined DNA Index System (CODIS) has already been described earlier. Since this is used to identify suspects and solve cases by comparing DNA profiles from crime scenes, it must be fool proof and highly accurate.

Purpose: The purpose of the development of CODIS is to facilitate the comparison of DNA profiles from crime scene evidence (like samples from rape kits or other evidence) with DNA profiles of convicted offenders and identify the suspect.

Software: The FBI provides the CODIS software, which enables public forensic laboratories to create searchable databases of authorized DNA profiles, as well as allowing them to share and compare DNA profiles through the NDIS.

How it Works: CODIS software allows state, local, and National law enforcement crime laboratories to compare DNA profiles electronically, linking crimes to each other and identifying potential suspects by matching DNA profiles from crime scenes with those from convicted offenders. Thus, CODIS is an extended invisible hand of law enforcement agents. CODIS database stores DNA profiles of convicts and other suspected individuals. This raises ethical questions.

How ethical is it to keep a database of convicted felons' DNA profiles? Can we rely on DNA fingerprints for conviction? Many ethical issues surround the use of DNA in forensic technology.

VIII. Nuances of Forensic DNA Fingerprinting

DNA is present in nearly every cell of our bodies, and we leave cells behind everywhere we go without even realizing it. Flakes of skin, drops of blood, hair, and saliva all contain DNA that can be used to identify us. The study of forensics, commonly used by police departments and prosecutors around the world, frequently relies upon these small bits of shed DNA to link criminals to the crimes they commit. This fascinating science is often portrayed on popular television shows as a simple, exact, and infallible method of finding a perpetrator and bringing him or her to justice. In truth, however, teasing out a DNA fingerprint and determining the likelihood of a match between a suspect and a crime scene is a complicated process that relies upon probability to a greater extent than most people realize. Government-administered DNA databases, such as the Combined DNA Index System (CODIS), do help speed the process, but they also bring to light complex ethical issues involving the rights of victims and suspects alike. Thus, far, we saw basics of DNA, allelic variation, DNA markers and

DNA fingerprinting. How this whole concept can be extended to uncover the allelic alteration in cancer cells is the main theme of this paper.

In earlier work (Afsheen *et al.,* 2025, in press) reported gradual evolution of DNA profiling technologies. DNA fingerprinting (Alec Jeffreys, 1985) was conducted using genomic DNA from the crime scene and that of the suspected individuals. The DNA after restriction digestion was resolved on agarose gel by electrophoresis. The DNA was then transferred onto the nylon membrane and hybridized with radio labeled genome derived cloned probe. After hybridization, the membrane was washed and exposed to X-ray file. This way the signal was recorded on the film (see Ali *et al.,* 1986; Afsheen *et al.,* 2025). In the process, DNA of the suspected person will match only when he was there at the scene of crime. Further, keeping in view the band profiles, one can statistically see the probability of two persons having identical band profiles. Now, DNA isolation and DNAF are automated using nucleic acids isolation machines and NGS along with multiple markers. A total of 16-20 STR markers are routinely used for forensic DNAF.

Next-generation sequencing (NGS) is a DNA sequencing technology that sequences multiple DNA fragments in parallel. NGS is also known as a massively parallel sequencing or deep sequencing system.

IX. Gradual Evolution of Sequencing Technology

Here we can see the gradual evolution of sequencing technology. This includes:

- First Generation
  - Sanger Sequencing which involves di-deoxy chain termination method.
- Second Generation Sequencing
  - Pyrosequencing
  - Sequencing by Reversible Terminator Chemistry
  - Sequencing by Ligation
- Third Generation Sequencing
  - Single Molecule Fluorescent Sequencing
  - Single Molecule Real Time Sequencing
  - Semiconductor Sequencing
  - Nanopore Sequencing
- Fourth Generation Sequencing
  - Aims conducting genomic analysis directly in the cell

NGS has several sub-types as discussed hereunder.

Table of Contents

- Next-Generation Sequencing Types
- Lynx therapeutics' massively parallel signature sequencing (MPSS)
- Polony sequencing
- Pyrosequencing
- Illumina (Solexa) sequencing
- SOLiD sequencing
- DNA nanoball sequencing
- Helioscope single molecule sequencing
- Single molecule SMRT sequencing
- Single molecule real time (RNAP) sequencing
- References

Our aim is to focus on allelic variation in cancer cells compared to that in normal ones. Highlighting all the available NGS sequencing nuances is not our aim. However, we will use NGS for allelic scores based on several well defined STR markers, discussion is imperative on NGS. Therefore, we provide a bird eye view on the nuances NGS.

X. Next Generation Sequencing (NGS) is a Robust platform That has Enabled the Sequencing of Thousands to Millions of DNA Molecules Simultaneously

Next-generation sequencing (NGS), also known as high-throughput sequencing, is the catch-all term used to describe a number of different modern sequencing technologies.

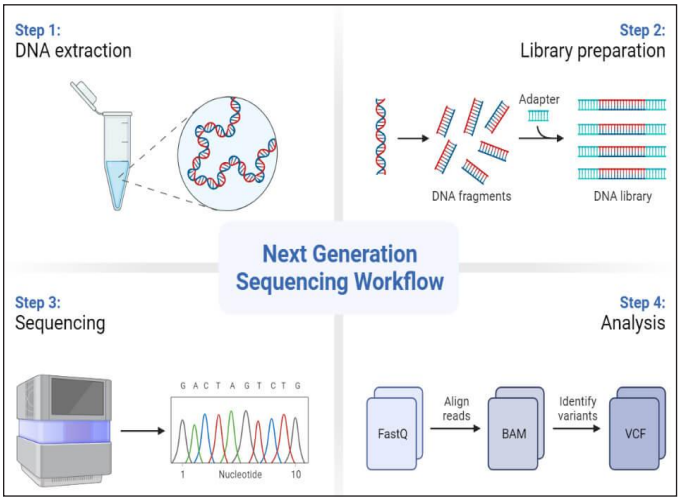


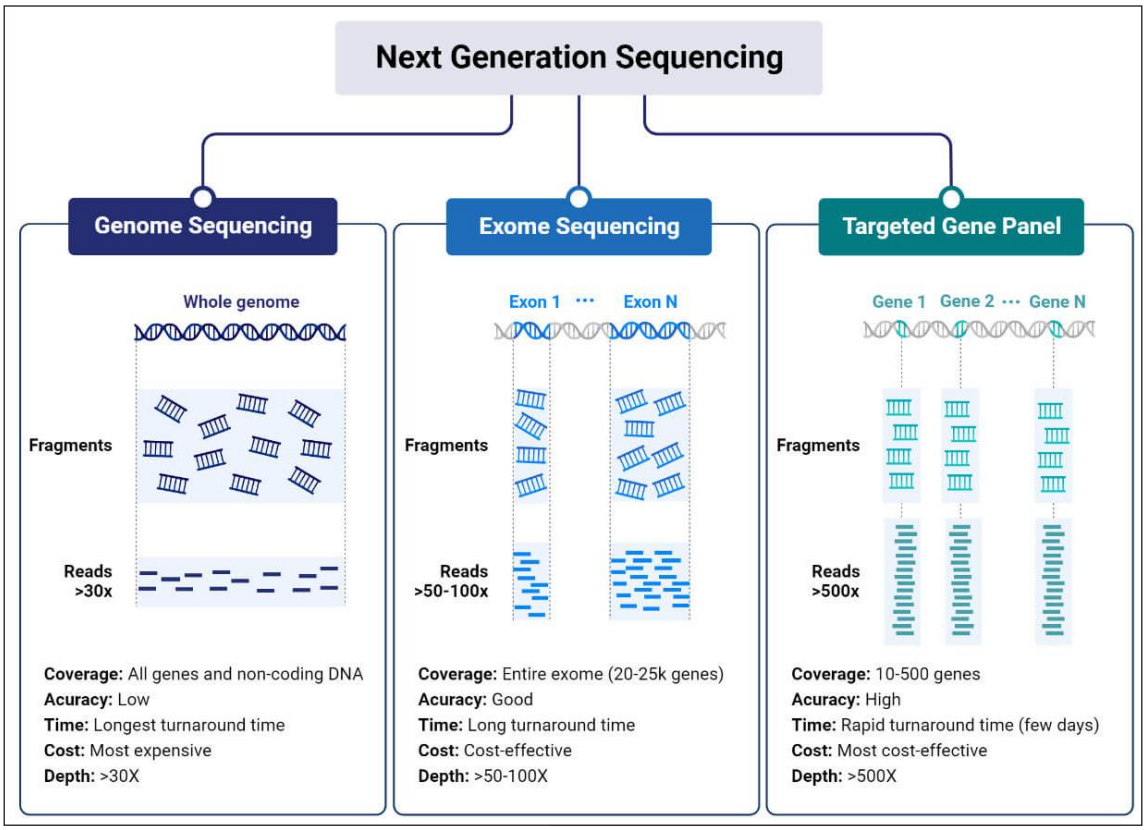
Fig. 1: This Diagrammatic Illustration Shows Next Generation Sequencing Workflow in Four Major Steps

The high demand for low-cost sequencing has driven the development of high-throughput sequencing, which produces thousands or millions of sequences at once. They

are intended to lower the cost of DNA sequencing beyond the standard dye-terminator methods. Thus, these recent technologies allow us to sequence DNA and RNA much more quickly and economically than the previously used Sanger sequencing (https://strbase-archive.nist.gov/tri_tab.htm). This has revolutionized the study of genomics and molecular biology. Classified to different generations, NGS has led to overcoming the limitations of conventional DNA sequencing methods and has found usage in a wide range of molecular biology applications.

XI. Brief Notes on the Working of NGS

- Isolate genetic material
- Prepare a library
- Sequence the DNA
- Analyse and interpret the data
- What it's used for
- Sequencing whole genomes
- Sequencing exomes
- Sequencing specific regions of interest
- Detecting variants and mutations
- Diagnosing genetic conditions like Alzheimer's disease, Parkinson's disease, and amyotrophic lateral sclerosis (ALS)
- Advantages
- NGS is faster and cheaper than previous sequencing technologies
- NGS can sequence hundreds or thousands of genes or an entire genome in a short period of time
- NGS is highly scalable and can be tuned to meet experimental needs
- NGS method includes Pyrosequencing, Sequencing by ligation (SOLiD), Sequencing by synthesis (SBS), and Ion Torrent sequencing. Following are the example of next generation sequencing.



Credit Bio -Render

Fig. 2: Examples of NGS Work Protocol

Here, we see another modified NGS platform though the basic principle remains the same (Fig. 3).

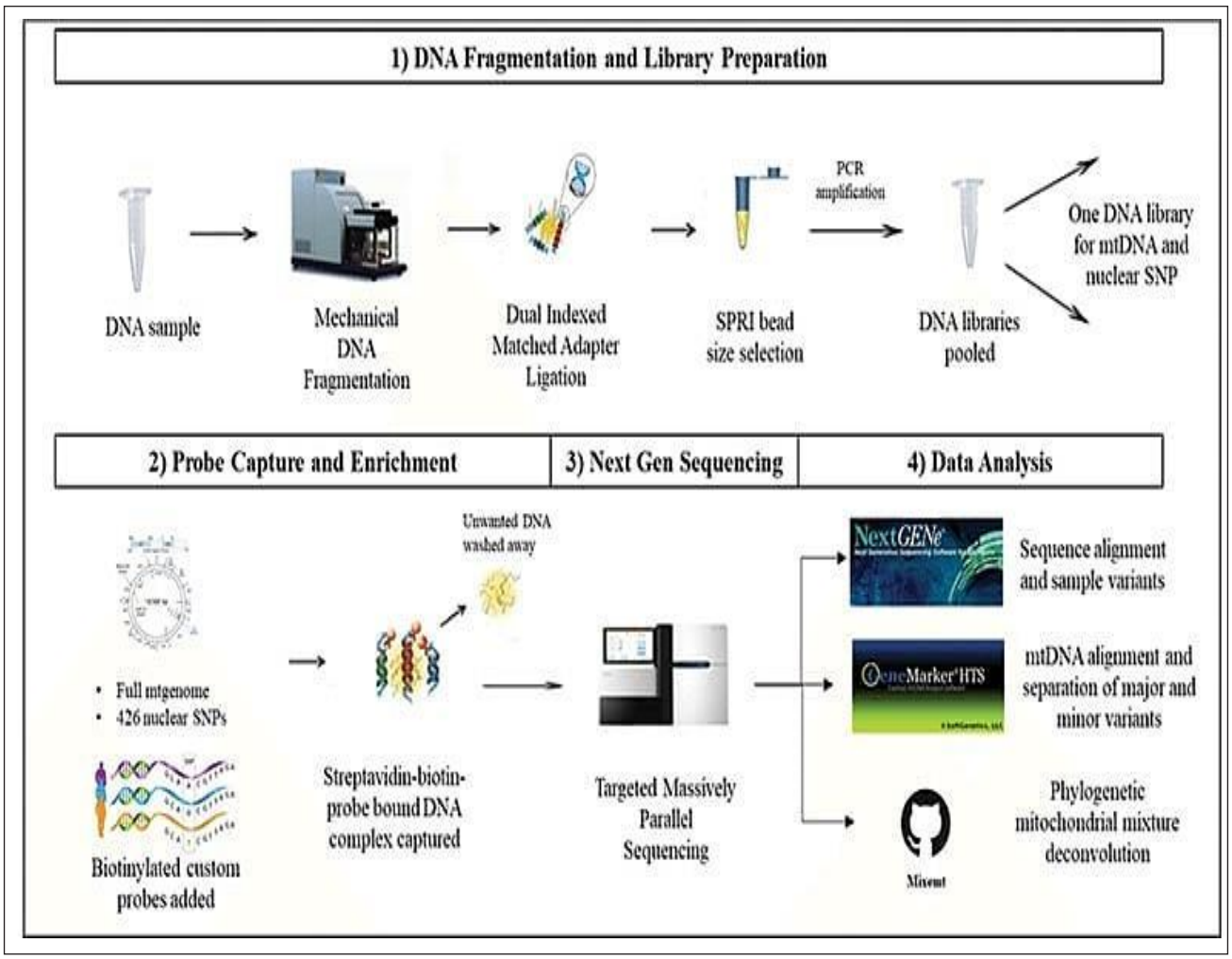


Fig. 3: Modified NGS Platform

Briefly, we will cover the following NGS types highlighting their salient features:

- Lynx therapeutics' massively parallel signature sequencing (MPSS)
- Polony sequencing
- Pyrosequencing
- Illumina (Solexa) sequencing
- SOLiD sequencing
- DNA nanoball sequencing
- Helioscope single molecule sequencing
- Single molecule SMRT sequencing
- Single molecule real time (RNAP) sequencing
- References

XII. Next-Generation Sequencing Types Lynx Therapeutics' Massively Parallel Signature Sequencing (MPSS)

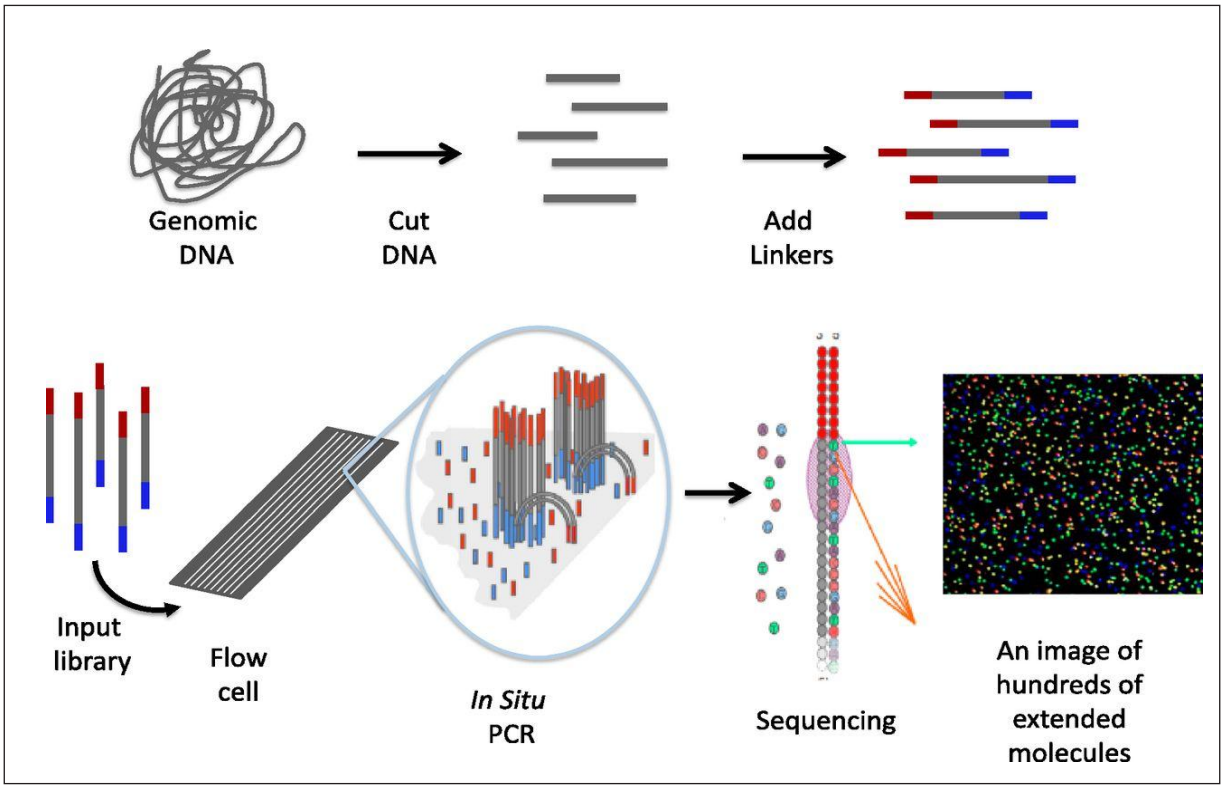


Fig. 4: NGS Types Lynx Therapeutics

Salient Features

- It is considered as the first of the "next-generation" sequencing technologies.
- MPSS was developed in the 1990s at Lynx Therapeutics, a company founded in 1992 by Sydney Brenner and Sam Eletr.
- Massively Parallel Signature Sequencing (MPSS) is an ultra-high throughput sequencing technology. This is a gene expression array technique that determines mRNA expression levels by counting individual mRNA molecules, using a method that combines in vitro cloning of DNA templates on micro beads with a ligation-based sequencing approach.
- When applied to expression profile, it reveals almost every transcript in the sample and provides an accurate expression level.
- MPSS was a bead-based method that used a complex approach of adapter ligation followed by adapter decoding, reading the sequence in increments of four nucleotides; this method made it susceptible to sequence-specific bias or loss of specific sequences.
- However, the essential properties of the MPSS output were typical of later "next-gen" data types, including hundreds of thousands of short DNA sequences.
- In the case of MPSS, these were typically used for sequencing cDNA for measurements of gene expression levels.

It may be noted that such a MPSS system is capable of handling large scale cDNA from different kinds of cancer tissue ensuring that each gene transcripts are not only made readily available but these can also be quantitated. Thus the copy number of a gene transcript from cancer cells and its comparison with those of normal ones would augment our understanding on cancer biology.

XIII. The Next NGS System is Polony Sequencing

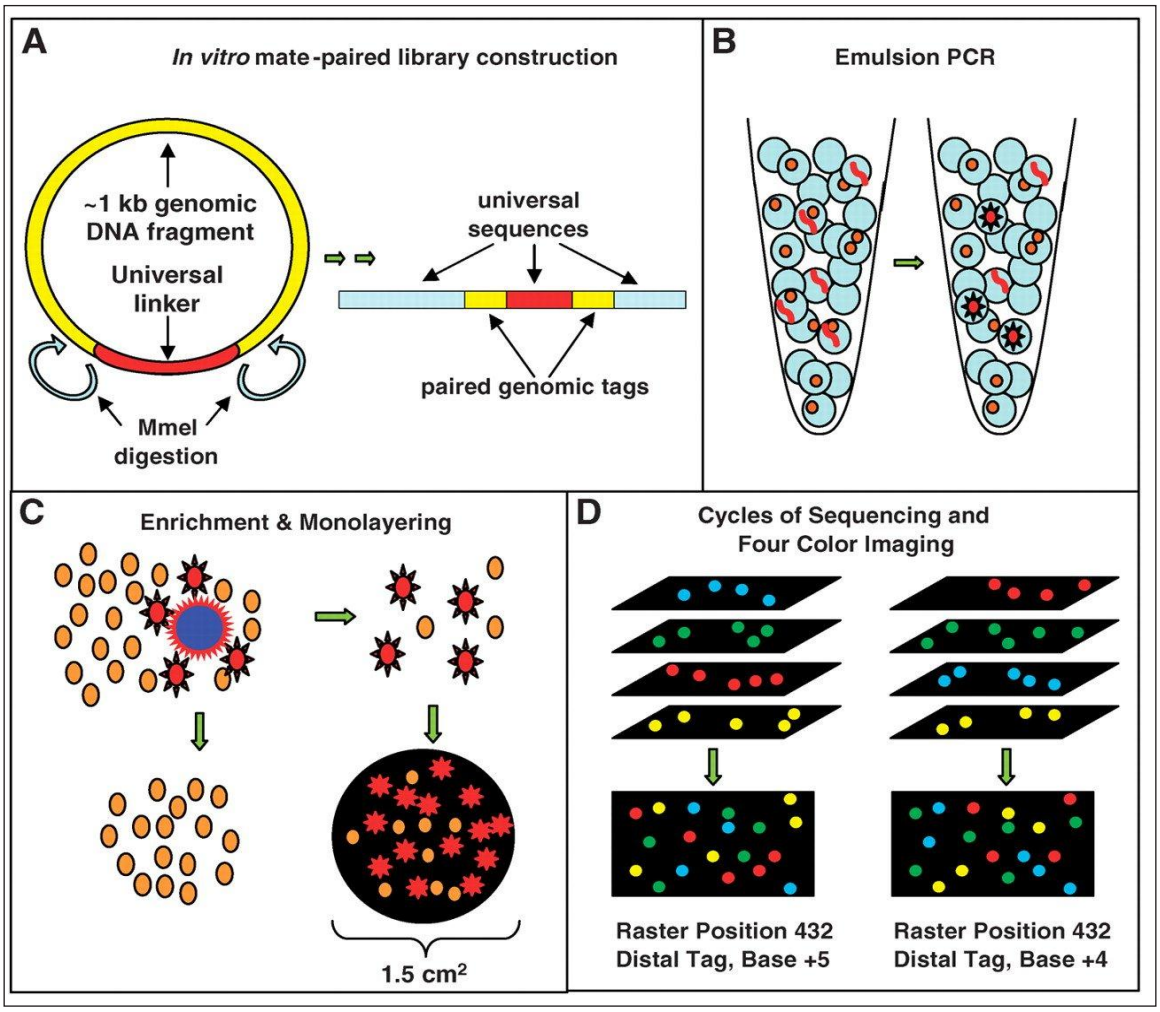


Fig. 5

Fig. 5, NGS Polony Sequencing showing the entire protocols in four simple steps.

Salient Features of Polony Sequencing

- It is an inexpensive but highly accurate multiplex sequencing technique that can be used to read millions of immobilized DNA sequences in parallel.
- This technique was first developed by Dr. George Church in Harvard Medical College.
- It combined an in vitro paired-tag library with emulsion PCR, an automated microscope, and ligation-based sequencing chemistry to sequence an *E. coli* genome at an accuracy of > 99.9999% and a cost approximately 1/10 that of Sanger sequencing.

XIV. Another Example of NGS is Pyrosequencing

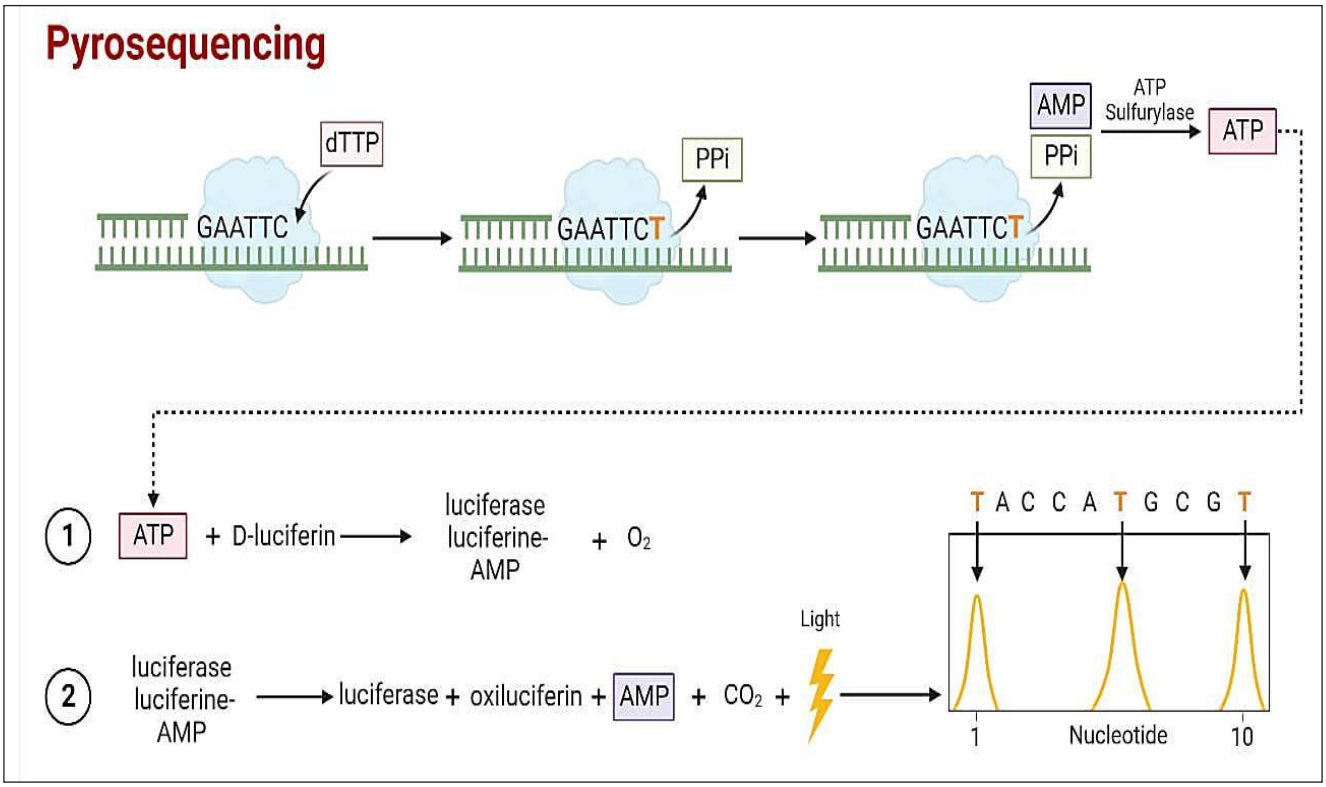


Fig. 6: NGS Pyrosequencing

The step by step protocols of this method is given in this diagrammatic illustration.

Salient Features of Pyrosequencing

- A parallelized version of pyrosequencing was developed by 454 Life Sciences, which has since been acquired by Roche Diagnostics.
- The method amplifies DNA inside water droplets in an oil solution (emulsion PCR), with each droplet containing a single DNA template attached to a single primer-coated bead that then forms a clonal colony.
- The sequencing machine contains many picolitre-volume wells each containing a single bead and sequencing enzymes.
- Pyrosequencing uses luciferase to generate light for detection of the individual nucleotides added to the nascent DNA, and the combined data are used to generate sequence read-outs.
- This technology provides intermediate read length and price per base compared to Sanger sequencing on one end and Solexa and SOLiD on the other.

XV. NGS Illumina (Solexa) Sequencing

We can now see Illumina (Solexa) Sequencing and its workflow shown in four simple steps.

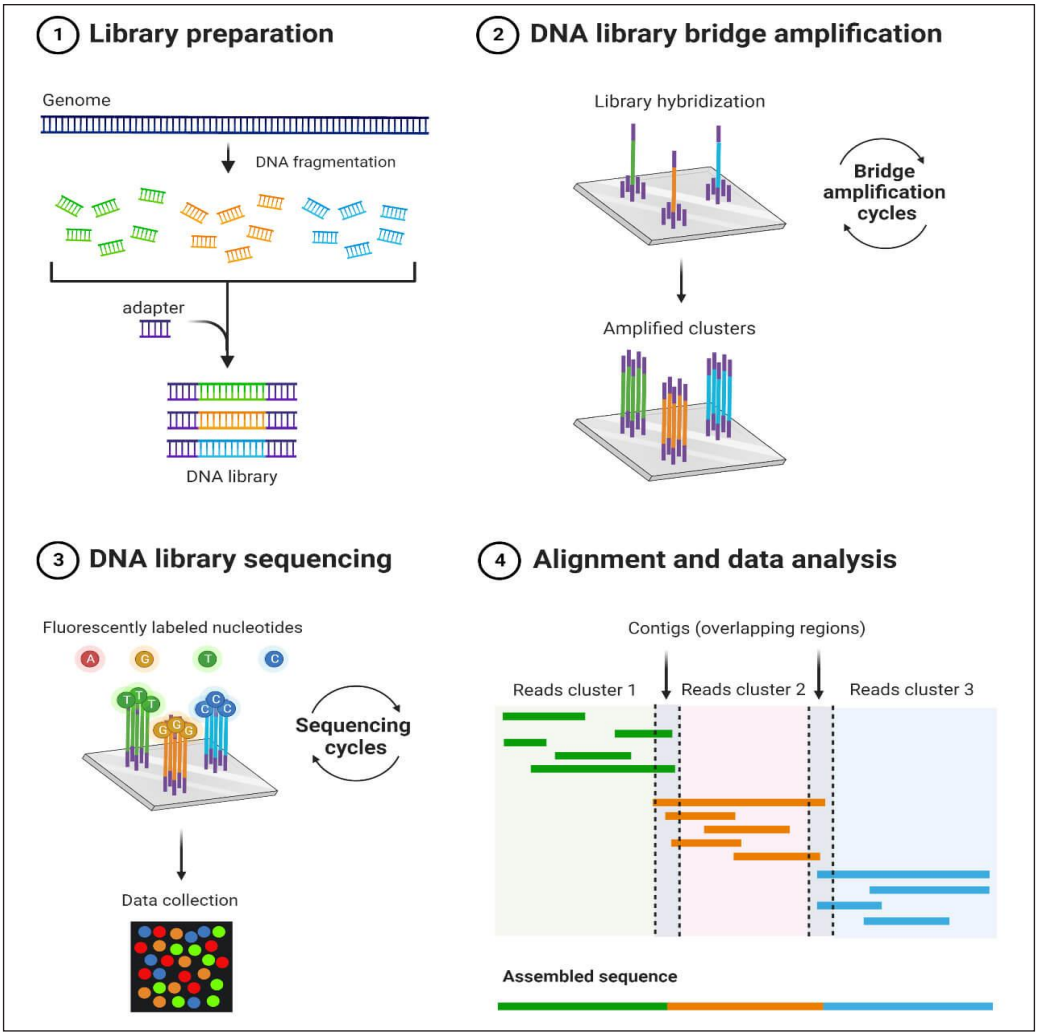


Fig. 7: NGS Illumina (Solexa)

Salient Features of llumina Sequencing

- Solexa developed a sequencing technology based on dye terminators.
- In this method, DNA molecule are first attached to primers on a slide and amplified. This is known as bridge amplification.
- Unlike pyrosequencing, the DNA can only be extended one nucleotide at a time.
- A camera takes images of the fluorescently labelled nucleotides, and then the dye along with the terminal 3′ blocker is chemically removed from the DNA, allowing the next cycle to commence.

XVI. NGS SOLiD Sequencing

Solid – Sequencing by Oligonucleotide Ligation and Detection is a next-generation DNA sequencing technology that uses ligation of dye-labelled oligonucleotides to determine DNA sequences known for its high accuracy.

SOLiD Sequencing: A Diagrammatic Illustration showing 8 steps of its operation.

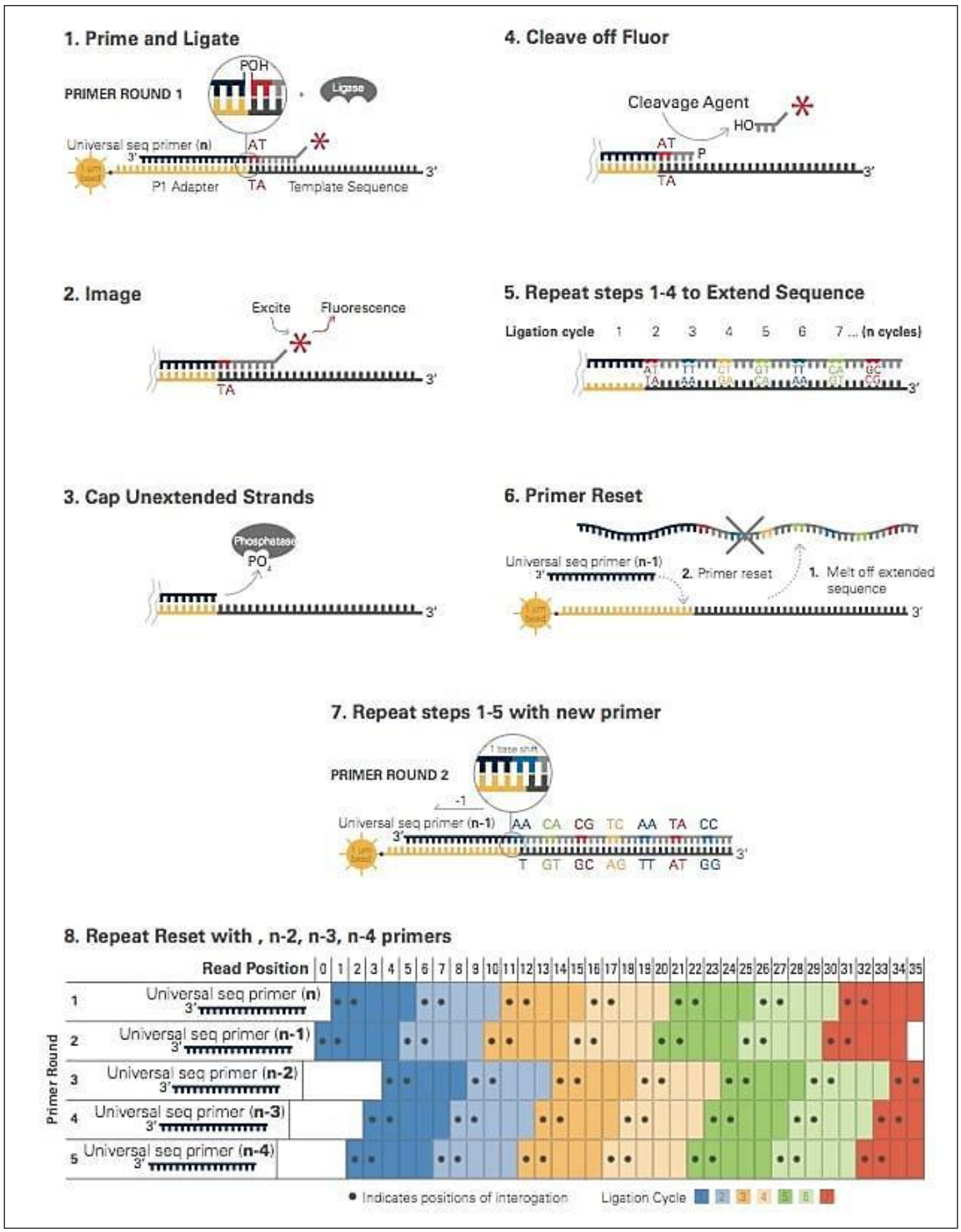


Fig. 8: NGS Solid Sequencing

Salient Features of SOLiD Sequencing

SOLiD is a next-generation sequencing (NGS) technology developed by Life Technologies (now Thermo Fisher Scientific) that is based on sequencing by ligation.

XVII. Workflow of SOLiD Sequencing

- *Ligation-Based:* The core principle is to sequence DNA by measuring the serial ligation of dye-labelled oligonucleotides to the DNA template.

- *Two-Base Encoding:* Each signal represents two bases in a row, rather than a single base, which contributes to the high accuracy of the method.
- *Emulsion PCR:* The DNA sample is amplified using emulsion PCR, where individual DNA molecules are amplified onto beads.
- *Ligation and Detection:* The beads are then used for SOLiD sequencing, where short, eight-base sequences (oligonucleotides) are used to form base pairing with the template DNA. The ligation of these probes is detected through a four-color imaging system.
- *Accuracy:* SOLiD is known for its high accuracy, with original base data accuracy greater than 99.94% and reaching 99.999% with a sequencing depth of 15x.
- *Applications:* SOLiD is suitable for various applications, including:
- *Targeted Sequencing:* Sequencing specific regions of the genome.
- *Resequencing:* Confirming high-throughput next-generation sequencing data.
- *Gene Panel Sequencing:* Sequencing specific sets of genes.

Advantages

High accuracy

Relatively inexpensive

Disadvantages

SOLiD sequencing is no longer commercially available.

XVIII. DNA Nanoball Sequencing

Following diagram shows DNA Nanoball sequencing workflow.

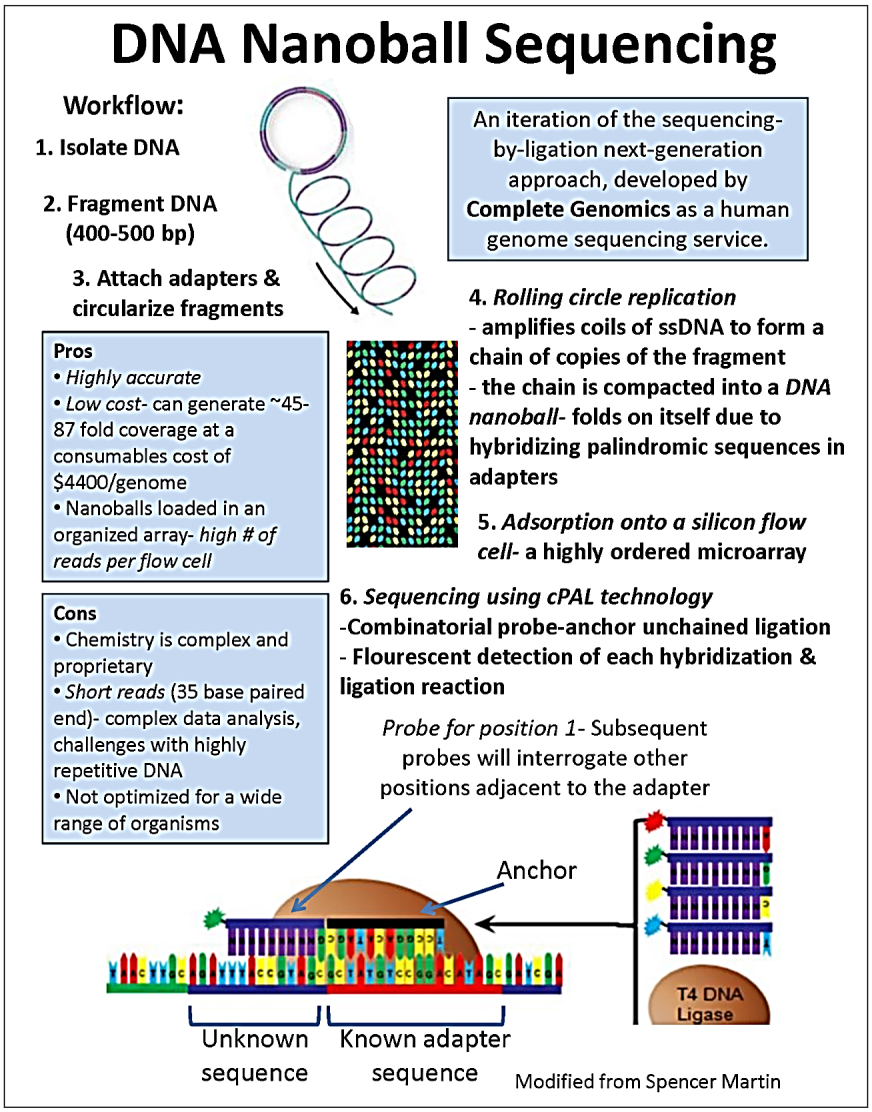


Fig. 9: DNA Nanoball Sequencing

XIX. Salient Features: DNA Nanoball Sequencing

- It is high throughput sequencing technology that is used to determine the entire genomic sequence of an organism.
- The method uses rolling circle replication to amplify fragments of genomic DNA molecules.
- This DNA sequencing allows large number of DNA nanoballs to be sequenced per run and at low reagent cost compared to other next generation sequencing platforms.
- However, only short sequences of DNA are determined from each DNA nanoball which makes mapping the short reads to a reference genome difficult.
- This technology has been used for multiple genome sequencing projects and is scheduled to be used for more.

XX. Helioscope Single Molecule Sequencing

Helioscope sequencing uses DNA fragments with added polyA tail adapters, which are attached to the flow cell surface. The next steps involve extension-based sequencing with cyclic washes of the flow cell with fluorescently labelled nucleotides. The reads are performed by the Helioscope sequencer. The reads are short, up to 55 bases per run, but recent improvement of the methodology allows more accurate reads of homopolymers and RNA sequencing.

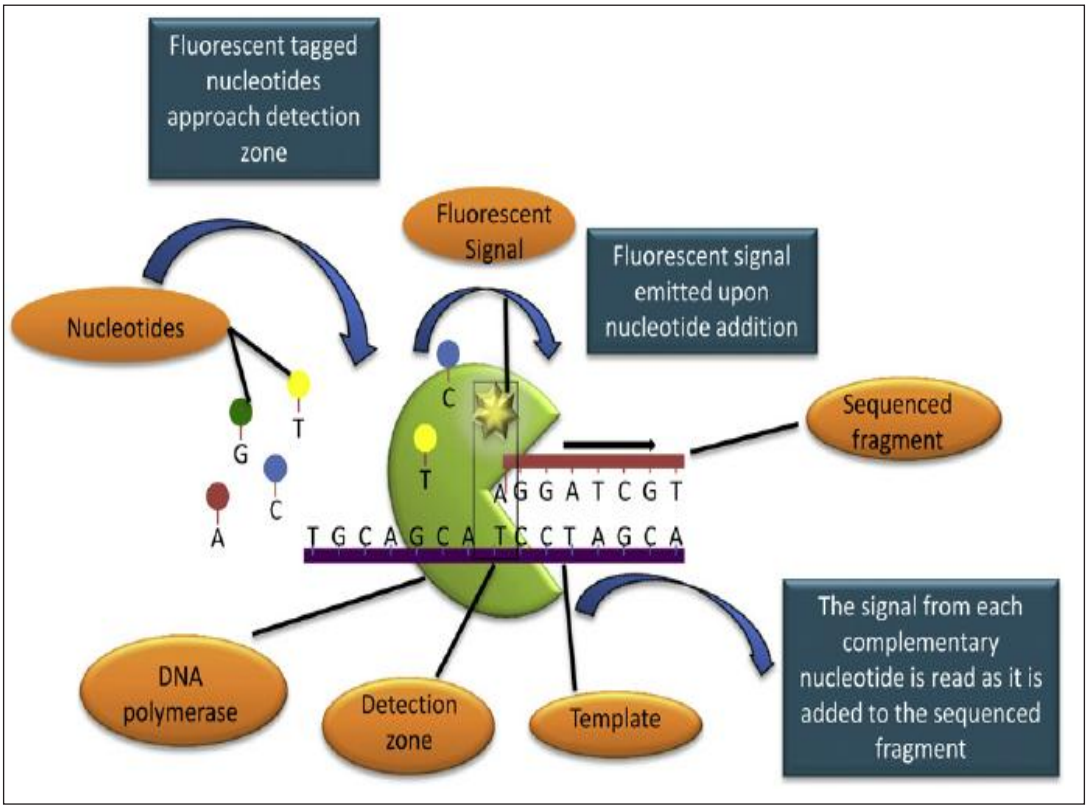


Fig. 10: Single Molecule SMRT Sequencing; Detailed Steps are Shown Here

Salient Features

- SMRT sequencing is based on the sequencing by synthesis approach.
- The DNA is synthesised in so called zero-mode wave-guides (ZMWs) – small well-like containers with the capturing tools located at the bottom of the well.
- The sequencing is performed with use of unmodified polymerase and fluorescently labelled nucleotides flowing freely in the solution.
- The wells are constructed in a way that only the fluorescence occurring by the bottom of the well is detected.
- The fluorescent label is detached from the nucleotide at its incorporation into the DNA strand, leaving an unmodified DNA strand.
- The SMTR technology allows detection of nucleotide modifications. This happens through the observation of polymerase kinetics.
- This approach allows reads of 1000 nucleotides.

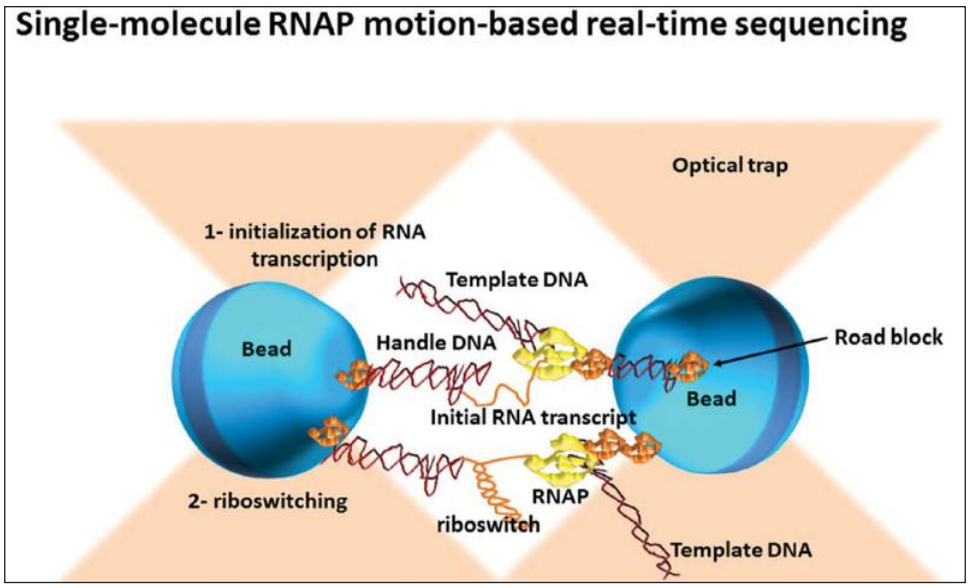


Fig. 11: Single Molecule Real Time (RNAP) Sequencing

Salient Features

- This method is based on RNA polymerase (RNAP), which is attached to a polystyrene bead, with distal end of sequenced DNA is attached to another bead, with both beads being placed in optical traps.
- RNAP motion during transcription brings the beads in closer and their relative distance changes, which can then be recorded at a single nucleotide resolution.
- The sequence is deduced based on the four readouts with lowered concentrations of each of the four nucleotide types.

We have seen 11 different NGS sequencing platforms from different companies. Each system has different methods for DNA sequencing with varying reads. However, with every method, fragments of DNA are amplified before the same is subjected to sequencing procedure. The NGS which is routinely used for human DNA fingerprinting globally is proposed to be used for this proposed project. Since a total of 64 STR markers are available, for 200 different types of cancers, a minimum of 12,800 reactions are envisaged. If genomic and cDNA both from cancer are used. This number will be just double totalling to be 27,600. Obviously, this involves seamless support and infrastructural facility. The ideal way is to check the hypothesis using established cancer cell lines and few STRs and then venture into this large scale academic game.

We have listed all the cancers and using this table one can select the promising one for this NGS based DNAF.

Earlier, it was largely believed that genetics and environment both have 50% roles each to play in causing cancer. However, current data suggests that genetics role causing cancer is only about 5-10%. It is the environment that seems to be the major cause (Fig. 12).

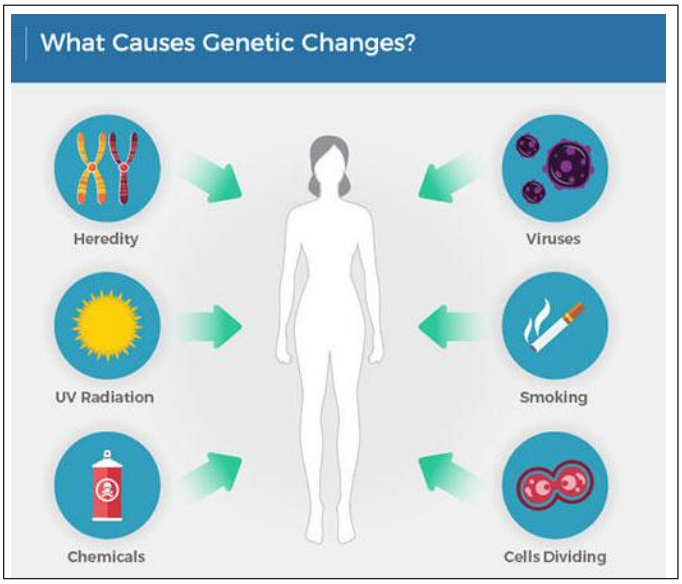
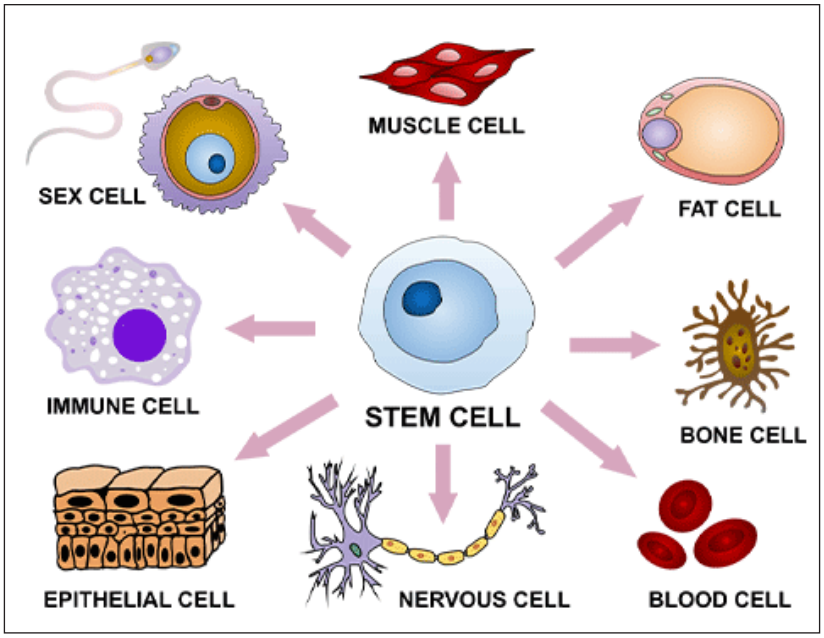


Fig. 12

Thus, Genetic changes that cause cancer can be inherited or arise from some environmental exposures. Genetic changes can also happen because of errors that occur during cell division (Credit: National Cancer Institute, NIH, USA).

When we look at the cells, we find them to be highly specialized with distinct morphology. When they become cancerous, they are deformed. Our aim is to record all such changes both at genotype and phenotype levels. As an example, some important normal cells with their morphology are shown in Fig. 13.



Source: https//common.wikimedia.org/wiki/File:Final_Stem_cell_differntiation_(1).svg

Fig. 13: Different Types of Normal Cells Showing Distinct Morphology

XXI. Different Types of Cells in Human Body

Life on earth deals with two kinds of cells. This includes prokaryotic and eukaryotic cells. Prokaryotic cells are primitive without membrane organelles and nuclei, whereas eukaryotic cells are advanced with membrane-bound nuclei and organelles. It has not been possible for humans to synthesize synthetic cytoplasm. This is because the cytoplasm is still mysterious. The human body is made up of about 200 types of eukaryotic cells. Cells come together to form tissues specified for a function and thus make organ systems. Based on different characteristics of cells in the human body, these can be divided into several types. Here we produce a table showing different types of cells based on morphology.

Table VII: Cell Types, Structures and Functions

Sr. No.	*Type of Cells*	*Structure*	*Function*
1	Platelets Cells	Small, irregularly shaped, nucleus absent	Blood clotting
2	Muscle Cells	Long, thin fibers	Contraction and movement
3	Nerve Cells	Star-shaped with long extensions	Carry information throughout the body
4	Germ Cells	Sperm cells small with head and tail; eggs are larger and spherical	Produce eggs and sperm
5	Red Blood Cells	Disc-shaped, no nucleus	Transport oxygen
6	White Blood Cells	Diverse shapes, defend against infection	Fight infection and disease
7	Stem Cells	Undifferentiated, can become many cell types	Self-renewal, differentiation into other cell types
8	Fat Cells	Large, spherical with single lipid droplet	Stores energy, insulates body, produces hormones
9	Skin Cells	Flattened, tightly packed with keratin	Protects body, provides barrier, detects sensation
10	Chondrocytes	Round, located in cartilage lacunae	Forms cartilage, provides support, absorbs shock

After fertilization, all the cells are differentiated into highly specialized ones. They tend to acquire newer mutations independently and thus their genotype will never be identical. This too will be reflected in the DNAF if appropriate STR markers are able to encompass those mutational regions. The human body is made up of 37 trillion cells. The nuances of these cells can be understood by studying cell biology and cytology. Fig. 14 shows how chromosomal rearrangement gives rise to fusion proteins. Chromosomes are studied using cell cytology.

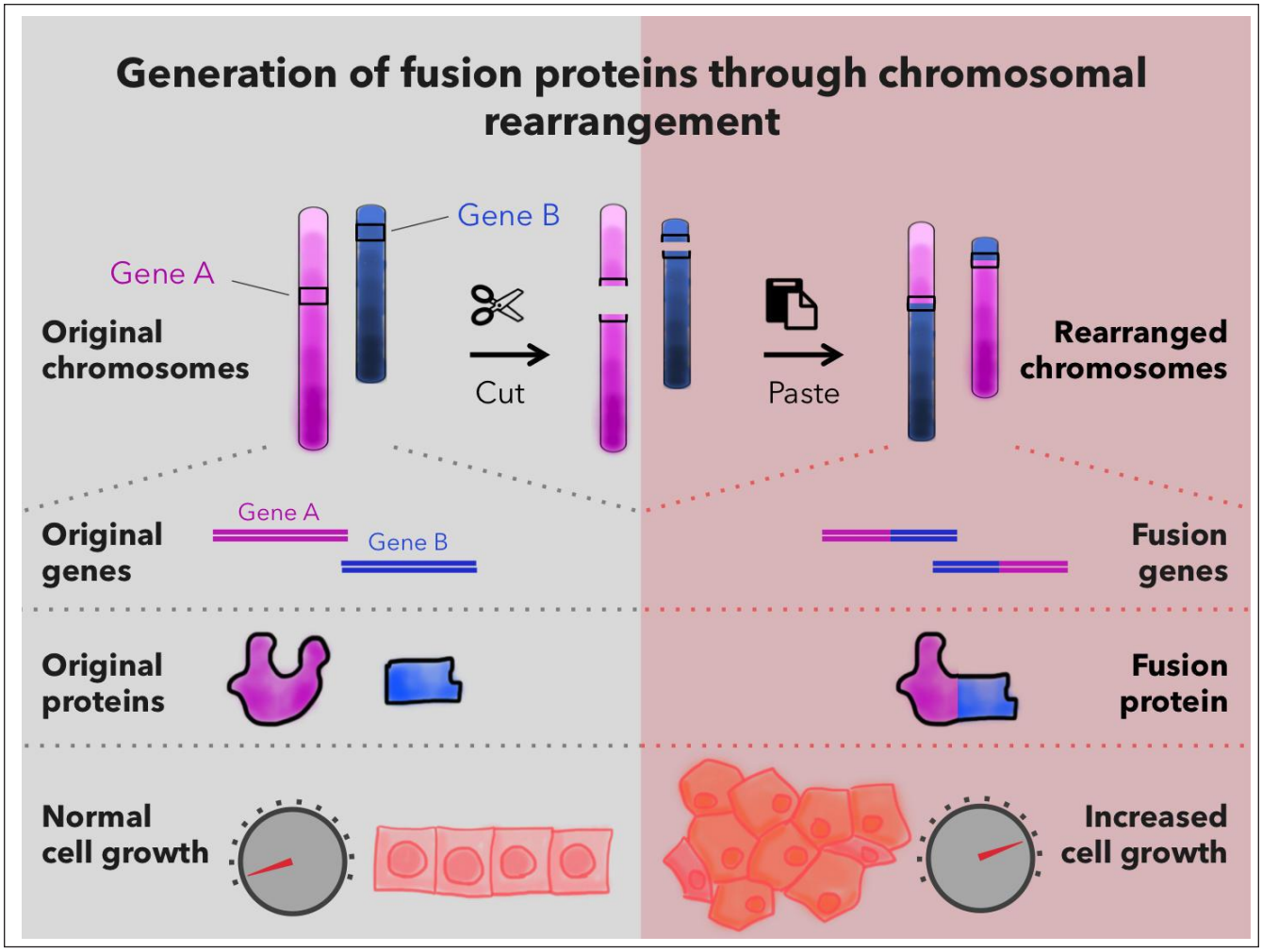


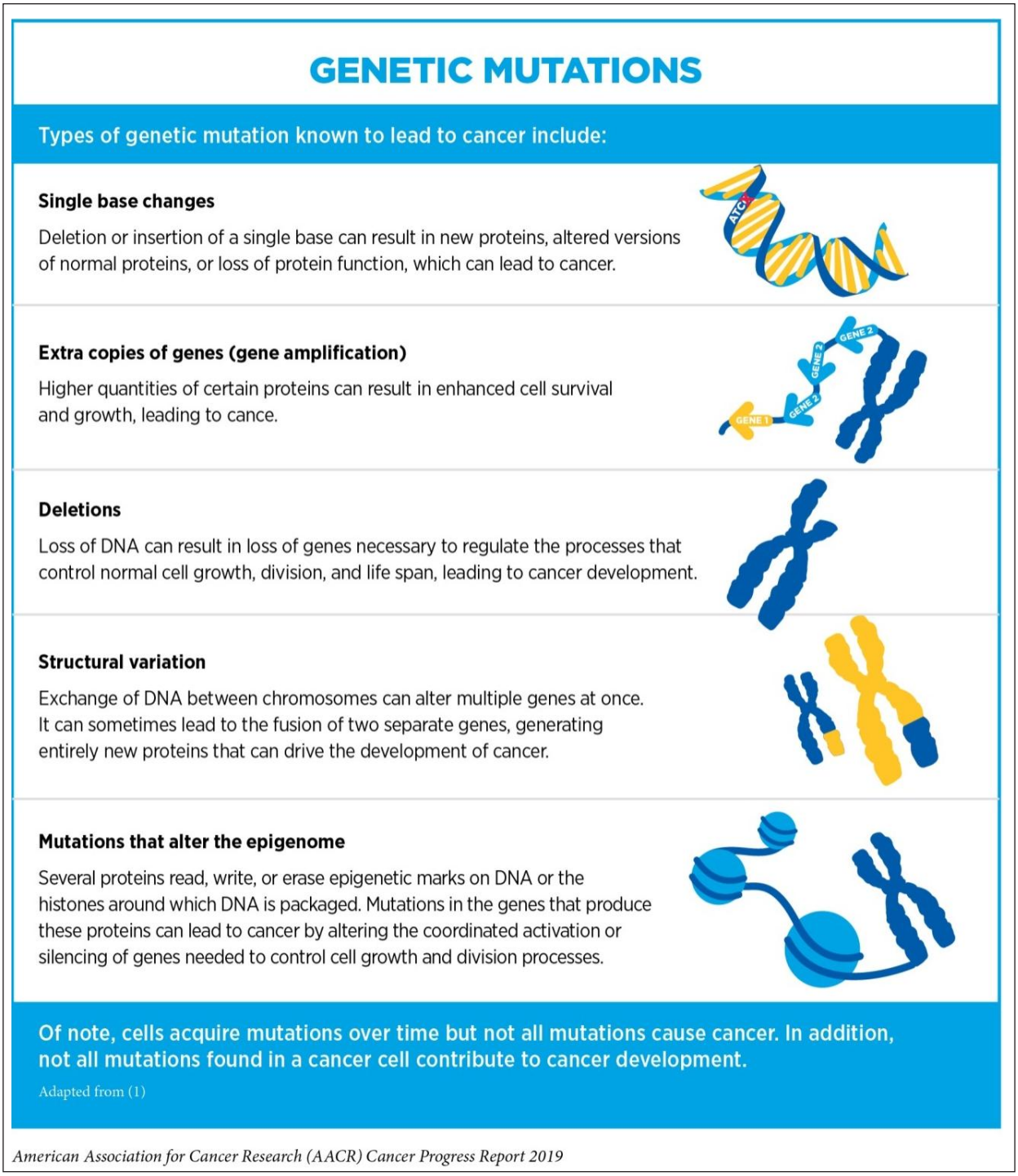
Fig. 14

Fig. 14, this figure shows generation of fusion proteins involving chromosomal rearrangement and its biological consequences.

The chromosomal rearrangement is one type of genetic alteration. There are other minor and major changes that take place in the genetic materials. Fig. 14 shows different types and levels of genetic mutations.

With the understanding of the genome organization, gene expression, mutation and chromosomal rearrangement, when we use the DNA for fingerprinting employing different well defined STRs, we find allelic profiles. These allelic profiles can be converted into numerical values. It becomes easy to compare two different DNA samples with respect to their similarity or dissimilarity. Allelic similarity will remain unchanged if the samples are the same despite the use of different STRs.

In order to explain the allelic profiles converted into numerical values and their comparison, we can see Table VIII. This table shows evidence samples and two suspects A and B. Suspect B has matching numerical values compared to evidence samples. In this table, we can see the use of 13 STR markers. However, the matching status suspect B remains unchanged clearly indicating that he is the culprit.



Credit: https://cancerprogressreport.aacr.org/

Fig. 15

Table VIII: Example of DNA Profiles Showing the STR Alleles for Each Sample and the Genotype Frequency of Suspect B for Each STR Locus

STR Locus	*Evidence Sample*	*Suspect A*	*Suspect B*	*Suspect B's Genotype Frequency for Each STR*
D3S1358	15, 17	17, 17	15, 17	0.13
vWA	15, 16	18, 19	15, 16	0.22
FGA	23, 27	21, 23	23, 27	0.31
D8S1179	12, 13	14, 15	12, 13	0.34
D21S11	28, 30	27, 30.2	28, 30	0.06
D18S51	12, 18	14, 18	12, 18	0.11
D5S818	13, 13	9, 12	13, 13	0.29
D13S317	12, 12	12, 12	12, 12	0.21
D7S820	10, 11	9, 10	10, 11	0.26
CSF1PO	8, 11	11, 12	8, 11	0.18
TPOX	7, 8	8, 8	7, 8	0.30
THO1	9.3, 9.3	6, 9.3	9.3, 9.3	0.38
D16S539	9, 13	11, 12	9, 13	0.10

XXII. Matching STRs Profiles Based on NGS Electropherograms

We have used this fictional table only to explain the concept STRs and matching of their profiles. These numerical values shown in the Table VII earlier are generated based on the electropherograms from the NGS sequencing Fig. 17.

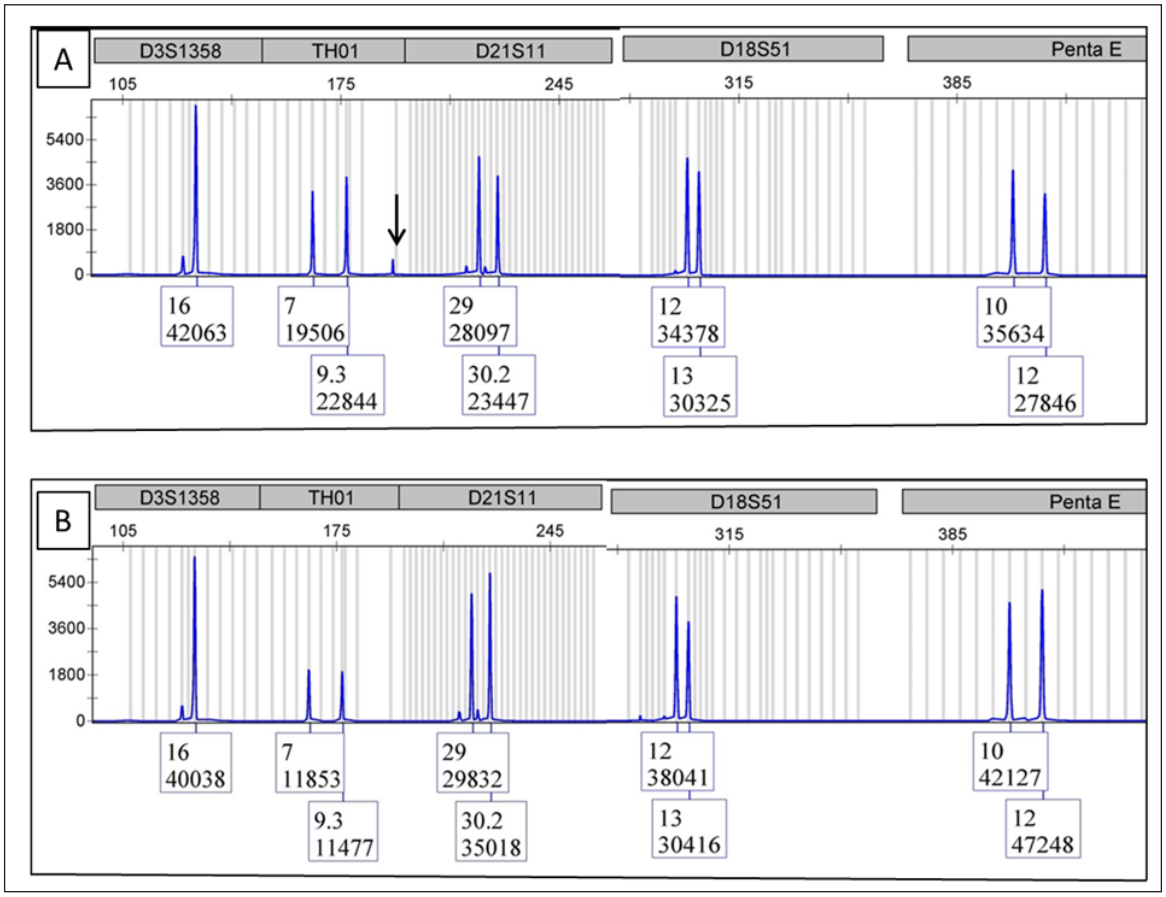


Fig. 16

Fig. 16, this electropherograms is based on the five different individual DNA samples showing similar allelic pattern of DNA fingerprint. The arrow in panel (A) indicates a non-specific peak. Note the similar profile in panel A and B. This data is generated using NGS sequencing platform.

This simple but most powerful concept may be extrapolated for cancer. If the cancer genotype is altered, the STR allelic profiles will be different compared to normal DNA samples. However, it is likely that one STR locus may not capture the altered alleles of the cancer cell. At this stage, use of multiple STR comes into picture. Fortunately, we have a total of 64 STRs though routinely only 16-20 are used. The cDNA from the cancer cells are expected to be more informative because gene expression is known to be tissue and stage specific. In the example shown in Table VIII, Suspect A is excluded as the source of the crime scene sample. Suspect B, on the other hand, matches the crime scene sample in all 13 STRs. A calculation of the frequency of Suspect B's genotype, based upon the STR allele frequencies within Suspect B's ethnic group, reveals that the likelihood that a random member of this ethnic group has this profile is about 1 in 1.5 billion. In view of this, when we subject cancer DNA to NGS sequencing based DNA fingerprinting, we expect an allelic profile, when we change the STR, the profile and its value will also change. However, when the same is compared with a normal DNA profile, the difference is likely to emerge as explained by the example in Table VIII.

XXIII. Concluding Remark

DNA fingerprinting technology has become well established and has been used successfully by law enforcement agent's world over. In addition, this approach has been equally useful in the areas of basic and biomedical research. Notwithstanding its enormous power, the same has not been tried to uncover the allelic profiles of cancer cells. When once work starts on this line data will start emerging and some part of the data will pose a challenge as well. It is envisaged that use of cancer cDNA from across the spectrum of cancer will prove to be highly enriching bringing new knowledge and newer understanding of right and wrong cell biology in the context of human health.

Acknowledgment

Authors are grateful to Professor Farzana Mahdi and Professor Abbas Ali Mahdi for their excellent administrative support and to Professor N B Singh for his continued academic indulgence.

References

[1] A. J. Jeffreys, V. Wilson, and S. L. Thein, "Individual specific 'fingerprints' of human DNA," *Nature,* vol. 316, 1985, pp. 76-79, 1985.

[2] S. Ali, C. R. Müller, and J. T. Epplen, "DNA fingerprinting by oligonucleotide probes specific for simple repeats," *Hum. Genet*., vol. 74, pp. 239-243, 1986.

[3] A. Fatima, F. Ali, and S. Ali, "Synthetic Oligodeoxyribonucleotides and human health," *Journal of Scientific and Technical Research (JSTR)*, 2025 in press.

[4] K. Norrgard, "Forensics, DNA fingerprinting, and CODIS," *Nature Education*, vol. 1, no. 1, p. 35, 2008.

[5] D. Pathak, S. Premi, J. Srivastava, S. Padinjarel Chandy, and S. Ali, "Genomic instability of the DYZ1 repeat in patients with Y chromosome anomalies and males exposed to natural background radiation," *DNA Research*, vol. 13, pp. 103-109, 2006.

[6] J. Srivastava, S. Premi, S. Kumar, and S. Ali, "Organization and differential expression of the GACA/GATA tagged somatic and spermatozoal transcriptomes in Buffalo *Bubalus bubalis," BMC Genomics,* vol. 9, pp. 132-148, 2008.

[7] J. Srivastava, S. Premi, S. Kumar, and S. Ali, "Expressional dynamics of minisatellite 33.15 tagged spermatozoal transcriptome in *Bubalus bubalis," BMC Genomics,* vol. 10, pp. 303-311, 2009.

[8] D. Pathak, and S. Ali, "RsaI repetitive DNA in buffalo *Bubalus bubalis* representing retrotransposons, conserved in bovids, are part of the functional genes," *BMC Genomics*, vol. 12, p. 338, 2011, doi: https://doi.org/10.1186/1471-2164-12-338.

[9] A. Bashamboo, and S. Ali, "Minisatellite associated sequence amplification (MASA) of the hypervariable repeat marker 33.15 reveals a male specific band in humans," *Mol Cell. Probes,* vol. 15, pp. 89-92, 2001.

[10] S. Ali, Md. A. Azfer, A. Bashamboo, P. K. Mathur, P. K. Malik, V. B. Mathur, A. K. Raha, and S. Ansari, "Characterization of a species specific repetitive DNA from a highly endangered wild animal *Rhinoceros unicornis* and assessment of genetic polymorphism by microsatellite associated sequence amplification (MASA)," *Gene,* vol. 4, no. 228, pp. 33-42, 1999. [This work was highlighted in Nature Biotechnology, Nov. 1999, vol. 16, p. 991].

[11] A. Bashamboo, and S. Ali, "DNA typing and micro-satellite associated sequence amplification (MASA) in detecting loss or gain of alleles," *Indian J. Biochem. Biophysics*, vol. 36, pp. 312-317, 1999.

[12] A. Kumari, S. K. Yadav, M. M. Misro, J. Ahmad, and S. Ali, "Copy number variation and microdeletions of

the Y chromosome linked genes and loci across different categories of Indian infertile males," *Scientific Reports,* vol. 5, p. 1778, 2025, doi: https://doi.org/10.1038/srep17738.

[13] D. Pathak, J. Srivastava, R. Samad, I. Parwez, S. Kumar, and S. Ali, "Genome wide search of genes tagged with the consensus of 33.6 repeat loci in buffalo *Bubalus bubalis* employing minisatellite associated sequence amplification," *Chromosome Res*, vol. 18, no. 4, pp. 441-458, 2010, doi: https://doi.org/10.1007/s10577-010-9132-0.

[14] S. Ali, and S. Gangadharan, "Differential evolution of coding and non-coding sequences in related vertebrates: Implications in probe design," *Proc. Ind. Nat. Sci. Acad. (PINSA)-B*66, pp. 49-67, 2000.

[15] C. Darwin, *The Origin of Species*. Published by United Kingdom of Great Britain and Ireland, 1859.

[16] R. Schëfer, S. Ali, and J. T. Epplen, "The organization of evolutionarily conserved GATA/GACA repeats in the mouse genome," *Chromosoma (Berl.)*, vol. 93, pp. 502-510, 1986.

[17] M. V. John, I. Parwez, M. V. S. Sivaram, S. Mehta, N. Marwah, and S. Ali, "Analysis of VNTR loci in fish genomes using synthetic oligodeoxyribonucleotide probes," *Gene,* vol. 172, pp. 191-197, 1996.

[18] P. Yam, L. D. Petz. S. Ali, D. Stock, and R. B. Wallace, "Development of a single probe for documentation of chimerism following bone marrow transplantation. *Am. J. Hum. Genet*., vol. 441, pp. 867-881, 1987.

[19] S. Ali, and R. B. Wallace, "Intrinsic polymorphism of variable number tandem repeat loci in the human genome," *Nucl. Acids Res*., vol. 16, pp. 8487-8496, 1988.

[20] M. Saini, A. N. Jha, R. Tangri, and Md. Q. Sher Ali, "MN1 overexpression with varying tumor grade is a promising predictor of survival of glioma patients," *Human Molecular Genetics*, vol. 29, no. 21, pp. 3532-3545, 2020, doi: https://doi.org/10.1093/hmg/ddaa231.