

# Το Πρόγραμμα Ανατομίας του Γονιδιώματος του Καρκίνου (CGAP). Μεθοδολογίες και κύρια Επιστημονικά Ευρήματα

Μαρίνα Χ. Καγκαρά

Βιολόγος Εκπαιδευτικός MSc ΚΦΕ/ΣΘΕΤ, ΕΑΠ  
kagaramarina@yahoo.gr

Ανδρέας Σκορίλας

Καθηγητής, Τομέας Βιοχημείας & Μορ. Βιολογίας,  
Τμήμα Βιολογίας, Πανεπιστήμιο Αθηνών.  
Μέλος ΣΕΠ / Συντονιστής ΚΦΕ53, ΚΦΕ/ΣΘΕΤ, ΕΑΠ  
ascorilas@biol.uoa.gr

**Περίληψη** – Στη συγκεκριμένη Μεταπτυχιακή Διπλωματική Εργασία μελετάται το Πρόγραμμα Ανατομίας του Γονιδιώματος του Καρκίνου (CGAP) και αναλύεται η συμβολή του στον τομέα της παθοβιολογίας, της πρόληψης, της έγκαιρης διάγνωσης και της αποτελεσματικής θεραπείας του καρκίνου. Ο ευρύς στόχος του προγράμματος ήταν η αποκρυπτογράφηση των γενετικών αλλαγών, που συμβαίνουν κατά τη διάρκεια της εξέλιξης του καρκίνου, και ο συνολικός μοριακός χαρακτηρισμός φυσιολογικών και καρκινικών κυττάρων. Στο πλαίσιο αυτό, δημιουργήθηκε ένας διαδικτυακός κατάλογος γονιδίων διαφόρων τύπων καρκίνου και θεμελιώθηκε η "πρωτοβουλία γονιδιακής περιγραφής" για την ανακάλυψη SNPs σε cDNA αλληλουχίες που σχετίζονται με τον καρκίνο. Επιπλέον υλοποιήθηκε λεπτομερής χαρακτηρισμός χρωμοσωμικών ανωμαλιών του καρκίνου. Τέλος, αναπτύχθηκαν διαδικτυακά λογισμικά εργαλεία για την *in silico* ανάλυση μεγάλων βιομοριακών συνόλων δεδομένων από τις EST και SAGE βάσεις δεδομένων.

**Λέξεις-κλειδιά:** CGAP, μη κωδικά RNAs, *in silico* ανάλυση γονιδιακής έκφρασης, καρκίνος, SAGE, ESTs

## I. ΕΙΣΑΓΩΓΗ

Μετά την ορόσημη περιγραφή της δομής της διπλής έλικας του DNA (1953), τη δημιουργία του πρώτου ανασυνδυασμένου μορίου DNA (1972), την ανάπτυξη μεθόδων αλληλούχισης DNA (1976), την καθιέρωση ανάπτυξης μεθόδων χαρτογράφησης του ανθρώπινου γονιδιώματος με βάση τους πολυμορφισμούς RFLPs και STSs (1980 & 1989) και τέλος με την ανάπτυξη της αλυσιδωτής αντίδρασης πολυμεράσης (1983), παρήχθη η κατάλληλη τεχνογνωσία για την επιτυχή χαρτογράφηση του ανθρώπινου γονιδιώματος από το Human Genome Project (Lander et al, 2001). Επισήμως το Πρόγραμμα ολοκληρώθηκε το 2003, εισάγοντας έτσι τη «Μετα-γονιδιωματική εποχή» (Baak et al, 2005). Η έρευνα για τον καρκίνο σε επίπεδο γονιδιώματος περιπλέκεται από το διαφορετικό, πολλές φορές, τρόπο γονιδιακής μεταγραφής και μετάφρασης σε λειτουργικά διαφορετικές πρωτεΐνες, οι οποίες ακολουθώντας υφίστανται πολλαπλές λειτουργικές αλλαγές (Baak et

al.,2005). Κάθε τύπος καρκίνου και κάθε στάδιο της εξέλιξής του έχει τη δική του μοριακή υπογραφή, η οποία δημιουργεί το φαινότυπο των αλλαγών που σημειώθηκαν στο κυτταρικό γονιδίωμα (Riggins & Strausberg, 2001, Gundem 2015).

Η μεγάλη εξέλιξη των τεχνολογιών μελέτης του γονιδιώματος και οι μελέτες του ανθρώπινου μεταγραφώματος αποκάλυψαν ότι περισσότερο από το 90% του ανθρώπινου γονιδιώματος μεταγράφεται, αλλά μόνο το 2% περίπου μεταφράζεται. Υπάρχει ένας εκπληκτικός αριθμός μη κωδικών RNAs (sncRNAs, mid-size ncRNAs, lncRNAs), τα οποία έχουν διαφορετικές βιολογικές λειτουργίες. Ο συνδυασμός των διαφόρων προσεγγίσεων έρευνας, σε επίπεδο γονιδιώματος, έχει αποτελέσει το έναυσμα για μια δραματική επανεκτίμηση του πληροφοριακού περιεχομένου του ανθρώπινου γονιδιώματος (Rinn et al., 2012, Gundem 2015).

Με αφετηρία τα τέλη της δεκαετίας του 1990, ένα μεγάλο μέρος των δεδομένων του γονιδιώματος και του μεταγραφώματος του καρκίνου παρέχονται από τα μεγάλης κλίμακας ερευνητικά προγράμματα όπως του *NCI's Cancer Genome Anatomy Project*, του *Sanger Centre's Cancer Genome Project* και του *Human Cancer Genome Project* του *Ludwig Institute for Cancer Research* (Strausberg et al, 2003). Τα προγράμματα αυτά συμπληρώνουν την αυξανόμενη γνώση συγκεκριμένων γενετικών μεταβολών στην έρευνα του καρκίνου, παρέχοντας μία θεμελιώδη βάση για μία εις βάθος έρευνα στην ογκογένεση (Strausberg & Simpson, 2010).

Κύριος σκοπός της συγκεκριμένης Μεταπτυχιακής Διπλωματικής εργασίας αποτέλεσε η διερεύνηση και η ανάλυση του Προγράμματος Ανατομίας του Γονιδιώματος του Καρκίνου, ώστε να δημιουργηθεί το κατάλληλο γνωστικό υπόβαθρο ως προς τις στρατηγικές και τις τεχνικές, που αναπτύχθηκαν και εφαρμόστηκαν στο CGAP, καθώς επίσης να επισημανθούν τα ευρήματα από τις αναλύσεις των προτύπων γονιδιακής έκφρασης σε φυσιολογικούς, προκαρκινικούς και καρκινικούς ιστούς.

## II. ΜΕΘΟΔΟΛΟΓΙΑ

Η πολυδιάστατη δομή του ερευνητικού προγράμματος CGAP κατέστησε αρχικά δύσκολη την βιβλιογραφική έρευνα, της οποίας σταθερός στόχος υπήρξε η επίτευξη μιας αξιόπιστης, ολοκληρωμένης και λεπτομερούς ανάλυσης των στρατηγικών του προγράμματος, των εφαρμοσμένων μοριακών και κυτταρογενετικών τεχνικών ανάλυσης υψηλής απόδοσης, καθώς επίσης και των επιμέρους νέων ερευνητικών προσεγγίσεων, που υιοθετήθηκαν στην πορεία (CCAP, SNP500Cancer), και της παράλληλης συνεργασίας του με άλλα ερευνητικά προγράμματα (Head και Neck CGAP, shRNA Clone Library και shRNA Validation Project). Με αυτά ως δεδομένα, για κάθε θεματική ενότητα της Μεταπτυχιακής Διπλωματικής εργασίας πραγματοποιήθηκε αρχική βιβλιογραφική αναζήτηση κυρίως σε βιβλία και άρθρα επιστημονικών περιοδικών του διεθνούς επιστημονικού ιστότοπου των βιοεπιστημών NCBI. Σημαντική πηγή άντλησης πληροφοριών σχετικά με τους διάφορους τομείς, τους στόχους, τα ευρήματα και τα διαδικτυακά λογισμικά εργαλεία του CGAP αποτέλεσε η αντίστοιχη ιστοσελίδα του <http://cgap.nci.nih.gov> και οι επιμέρους ενότητές της. Παράλληλα, μελετήθηκαν βάσεις δεδομένων αλληλουχιών dbEST και UniGene του NCBI που περιέχουν αλληλουχίες ESTs (Expressed Sequence Tags, Ετικέτες Εκφραζόμενων Αλληλουχιών) φυσιολογικών και καρκινικών κυττάρων του CGAP, καθώς και βάσεις δεδομένων αλληλουχιών SAGE του CGAP (SAGEmap). Άλλες βάσεις δεδομένων που διερευνήθηκαν για δεδομένα συσχέτισης ιστοειδικών προτύπων γονιδιακής έκφρασης και των πολυμορφισμών που προσδιορίζονται σε διάφορα χρωμοσώματα είναι οι: i) Cancer Genome Atlas (TCGA) <http://cancergenome.nih.gov/>, ii) The Cancer Gene Census και Catalogue Of Somatic Mutations In Cancer <http://cancer.sanger.ac.uk/cosmic/> και iii) η βάση δεδομένων χρωμοσωμικών αλλοιώσεων Mitelman (Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer), <http://cgap.nci.nih.gov/Chromosomes/Mitelman>, η οποία περιέχει όλους τους δημοσιευμένους καρύοτυπους νεοπλασιών και τις αντίστοιχες γνωστές μεταλλαγές.

## III. ΑΠΟΤΕΛΕΣΜΑΤΑ

Στο πλαίσιο της συγκεκριμένης Μεταπτυχιακής Διπλωματικής εργασίας, από την ανάλυση των μεθοδολογιών και των κύριων ευρημάτων του CGAP, δείχθηκε ότι έχει αναπτυχθεί μία κεντρική αφετηρία

για σημαντικά μοριακά εργαλεία. Συγκεκριμένα, με την δημιουργία της αντίστοιχης ιστοσελίδας, το σύνολο των επιτευγμάτων του έργου γίνεται άμεσα προσβάσιμο στην ευρεία επιστημονική κοινότητα έρευνας του καρκίνου χωρίς περιορισμούς. Καθιστώντας διαθέσιμες διαδικτυακά τις μεταγραφόμενες αλληλουχίες του ανθρώπου και του ποντικού, συνέβαλε στην ραγδαία επιτάχυνση της έρευνας του γονιδιώματος του καρκίνου αλλά και άλλων ερευνών στον τομέα της Μοριακής Βιολογίας. Οι βάσεις δεδομένων αλληλουχιών CGAP και η αντίστοιχη ιστοσελίδα <http://cgap.nci.nih.gov> συγκροτεί, σε ένα ενιαίο σύνολο, ποικίλες πηγές δεδομένων (προτύπων γονιδιακής έκφρασης, χρωμοσωμικών ανωμαλιών και πολυμορφισμών, σχολιασμούς λειτουργικότητας γονιδίων, βιοχημικά μονοπάτια σηματοδότησης) και λογισμικά εργαλεία ανάλυσης, επιτρέποντας στον ερευνητή να βρει "*in silico*" απαντήσεις σε βιολογικά ερωτήματα σε μικρό χρονικό διάστημα [Riggins & Strausberg, 2001, Gundem 2015]. Η εφαρμογή των εργαλείων της πληροφορικής του CGAP, σε συνδυασμό με εργαστηριακά πειράματα, συνέβαλλαν και θα συνεχίσουν να συμβάλλουν στον εντοπισμό πολλών άγνωστων γονιδίων, που εκφράζονται σε ανθρώπινους ιστούς (Strausberg et al., 2000). Επιπρόσθετα, η χρησιμότητα της βάσης δεδομένων CGAP τεκμηριώνεται από την παγκόσμια ερευνητική κοινότητα στη λήψη δεδομένων για πειραματικές διαδικασίες. Πολυάριθμες δημοσιεύσεις βασίζονται στη χρήση των βάσεων δεδομένων του CGAP για την ταυτοποίηση νέων γονιδίων του καρκίνου, των οποίων τα προϊόντα υπερεκφράζονται ή υποεκφράζονται σε διάφορους τύπους καρκίνου, συμπεριλαμβανομένων του προστάτη, του μαστού, του παχέος εντέρου, των ωοθηκών, του πνεύμονα, του παγκρέατος και του εγκεφάλου. Οι στρατηγικές λήψης δεδομένων και οι ακόλουθες αναλύσεις διαφέρουν, αντανακλώντας την δημιουργικότητα που μπορεί να εφαρμοστεί στη βάση δεδομένων (Strausberg et al., 2001). Στο πλαίσιο του προγράμματος CGAP ιδρύθηκε η βάση δεδομένων *Tumor Gene Index* (TGI) η οποία θα χρησιμεύσει ως ένας κατάλογος όλων των γονιδίων που εκφράζονται σε ολόκληρο το φάσμα της εξέλιξης του καρκίνου, με έμφαση στους καρκίνους του προστάτη, του μαστού, των ωοθηκών, του πνεύμονα και του παχέος εντέρου και με ιδιαίτερη αναφορά στον τύπο του όγκου και στο στάδιο της εξέλιξης. Επέκταση αυτού αποτέλεσε η ίδρυση αντίστοιχου γονιδιακού ευρετηρίου όγκων (*Tumor Gene Index*, TGI) σε οργανισμούς μοντέλα, όπως του ποντικού, από τους παραπάνω πέντε τύπους καρκίνου. Στο TGI τα γονίδια θα σχολιάζονται ως προς τον τύπο της cDNA βιβλιοθήκης, τον ιστό προέλευσης, το στάδιο του καρκίνου και την

εξειδίκευση της γονιδιακής έκφρασης (Krizman et. al, 1999; Strausberg et al., 1997; Strausberg et al, 2001; Strausberg et al, 2001; Strausberg et. al., 2000).

Μια σημαντική χρήση των καταλόγων γονιδίων του CGAP είναι ο σχεδιασμός νέων μικροσυστοιχιών DNA για την εύρεση και μελέτη πρότυπων της γονιδιακής έκφρασης χιλιάδων γονιδίων, σε μεγάλα σύνολα ιστών με άξονα τα διαφορετικά στάδια της καρκινογένεσης (Aldaz M., 2003; Porter et.al, 2001). Η γνώση αυτών των εκφρασμένων γονιδίων μπορεί στη συνέχεια να εφαρμοστεί, για να καθορίσει μοριακές υπογραφές καρκίνων, ώστε να χρησιμοποιηθούν ως βιοδείκτες για την έγκαιρη διάγνωση την πρόγνωση και την επιλογή της καταλληλότερης θεραπευτικής αγωγής (Riggins & Strausberg, 2001; Strausberg et al., 1997).

Το πρότυπο των αλλαγών της γονιδιακής έκφρασης που συμβαίνουν, καθώς ένα φυσιολογικό κύτταρο μετασχηματίζεται σε καρκινικό, μπορεί να παρέχει μία εικόνα για τις μοριακές αλλαγές που συμβαίνουν κατά την εξέλιξη της καρκινογένεσης. Έτσι, το CGAP αρχικά υιοθέτησε δύο κύριες στρατηγικές για τον προσδιορισμό των προτύπων γονιδιακής έκφρασης: α) τη χρήση των ESTs από τη βάση δεδομένων dbEST, καθιστώντας δυνατή την ταυτοποίηση ιστοειδικών γονιδίων, αφού η παρουσία ή η απουσία των ESTs σε διαφορετικές βιβλιοθήκες παρέχει πληροφορίες σχετικά με το όργανο, τον κυτταρικό τύπο ή την εξειδίκευση των εκφρασμένων γονιδίων ως προς το είδος του όγκου (Strausberg & Riggins, 2001; Riggins & Strausberg, 2001; Vasmatazis et al., 1998) και β) τη χρήση των SAGE αλληλουχιών από την βάση δεδομένων SAGEmap, την οποία δημιούργησε το CGAP σε συνεργασία με το NCBI. Η SAGE είναι μία μέθοδος που μετρά ποσοτικά τα επίπεδα μεταγραφής προσδιορίζοντας μεγάλο πλήθος μετάγραφα αλλά και σπάνια μετάγραφα. Το λογισμικό εργαλείο SAGE Genie δημιουργήθηκε από το CGAP, για την βέλτιστη ανάλυση και διάδοση αυτών των ψηφιακών προφίλ γονιδιακής έκφρασης (Strausberg RL., 2001; Lash et al., 2000; Lal et.al, 1999). Με την εύρεση λεπτομερών προτύπων γονιδιακής έκφρασης από καρκινικούς και φυσιολογικούς κυτταρικούς τύπους, προσφέρει ένα αυξανόμενης ισχύος εργαλείο στη διάγνωση, πρόγνωση και θεραπεία του καρκίνου (Strausberg & Riggins, 2001; Riggins & Strausberg, 2001, Gundem 2015). Επιπλέον, βιβλιοθήκες σχεδιασμένες για τον προσδιορισμό των γονιδίων που εμπλέκονται στην ανοχή φαρμάκων έχουν πρόσφατα προστεθεί στο πρόγραμμα μέσω του SAGEmap (Strausberg & Riggins, 2001; Boon et al., 2002; Zhang et. al., 2015).

Τα αποτελέσματα της συνεργασίας του CGAP και του προγράμματος *Head and Neck Cancer Genome Anatomy Project* (HN-CGAP) αφορούν την συστηματική ταυτοποίηση και καταγραφή των

γνωστών και νέων γονιδίων, που εκφράζονται κατά τη διάρκεια της ανάπτυξης του καρκινώματος εκ πλακωδών κυττάρων της κεφαλής και του τραχήλου (Head και Neck Squamous Cell Cancer, HNSCC), με τη χρήση της τεχνολογίας των DNA συστοιχιών. Η ανάλυση των προτύπων γονιδιακής έκφρασης θα συμβάλει στον ολοκληρωμένο μοριακό χαρακτηρισμό των φυσιολογικών, προκαρκινικών και καρκινικών κυττάρων και τελικά στην διαλεύκανση των άγνωστων μηχανισμών που εμπλέκονται στην καρκινογένεση. Αναμένεται, επίσης, το μοριακό πρότυπο για τον HNSCC να οδηγήσει στην εύρεση κατάλληλων βιοδεικτών για την έγκαιρη ανίχνευση των προνεοπλαστικών αλλοιώσεων και στον εντοπισμό των γονιδιακών προϊόντων που μπορούν να καθορίσουν το μετασχηματισμένο και μεταστατικό φαινότυπο καθώς και στον σχεδιασμό νέων θεραπευτικών προσεγγίσεων (Patel&Leethanakul, 2001; Packer et al, 2006; Baak1 et al, 2005, Gundem 2015). Προς αυτή την κατεύθυνση στο πρόγραμμα SNP500Cancer, το οποίο αποτελεί μέρος του CGAP, στοχεύει στην ταυτοποίηση και τον χαρακτηρισμό της γενετικής ποικιλότητας σε γονίδια σημαντικά στον καρκίνο καθώς και στην ανακάλυψη νέων SNPs και άλλων σημαντικών γενετικών παραλλαγών δυναμικής σημασίας για τις μελέτες της μοριακής επιδημιολογίας του καρκίνου. Οι πολυμορφισμοί ενός νουκλεοτιδίου (SNPs), που έχουν επικυρωθεί σε αυτήν την πρωτοβουλία, χρησιμοποιούνται από αρκετά ερευνητικά κέντρα για τη γονοτύπηση καρκινικών δειγμάτων και για τις επιδημιολογικές έρευνες σχετικά με τους περιβαλλοντικούς και γενετικούς παράγοντες του καρκίνου (Packer. et al, 2004; Packer et al, 2006). Στο πλαίσιο των αποτελεσμάτων της εργασίας αναλύεται, επίσης, η συμβολή του CGAP στην προώθηση της τεχνολογίας παρεμβολής RNA (RNA interference, RNAi) μέσω της συνεργασίας του με άλλα ερευνητικά προγράμματα όπως τα shRNA Clone Library και shRNA Validation Project που οδήγησαν στην κατασκευή shRNA κλώνων και στο σχεδιασμό μια σειράς από shRNAs που κατατέθηκαν σε δημόσια ηλεκτρονική βάση δεδομένων. Όλα τα παραγόμενα shRNAs μπορούν να χρησιμεύσουν ως εργαλεία στόχευσης και ρύθμισης της έκφρασης γνωστών γονιδίων που σχετίζονται με τον καρκίνο (<http://cgap.nci.nih.gov/RNAi>).

#### IV. ΣΥΜΠΕΡΑΣΜΑΤΑ

Σταθερός στόχος, κατά την εκπόνηση της συγκεκριμένης Μεταπτυχιακής Διπλωματικής εργασίας, υπήρξε πρωτίστως η επισήμανση της συμβολής του ερευνητικού προγράμματος CGAP στην έρευνα του καρκίνου, μέσω της ανάδειξής του ως ένα

αποτελεσματικό εργαλείο Μοριακής Βιολογίας της επιστημονικής κοινότητας. Η εργασία αυτή συνιστά μία μοναδική ελληνική πηγή άντλησης δεδομένων των διαφόρων πτυχών του προγράμματος CGAP και με τη συλλογή των σχετικών βιβλιογραφικών αναφορών μπορεί να συμβάλει στην ενημέρωση των βιοεπιστημόνων. Υπό αυτή τη θεώρηση, διευκολύνεται η αξιοποίηση των δεδομένων του Προγράμματος σε μελλοντικές πειραματικές διαδικασίες και σε *in silico* αναλύσεις στην έρευνα του καρκίνου.

Εξίσου σημαντική είναι η προσδοκία της συμβολής της εργασίας αυτής πρωταρχικώς στην επιμόρφωση των εκπαιδευτικών των Φυσικών Επιστημών της Δευτεροβάθμιας Εκπαίδευσης και ακολούθως, μέσω της εκλαϊκευμένης θεματικής της ενότητας, στην αποδοτικότερη επαφή των μαθητών με τις εξελίξεις στην παγκόσμια έρευνα του καρκίνου, καθώς επίσης και με τις εφαρμογές των διαφόρων επιστημονικών ευρημάτων στην πρόληψη, διάγνωση, πρόγνωση και θεραπεία του καρκίνου. Η ενημέρωση φοιτητών και μαθητών σχετικά με το γονιδίωμα του καρκίνου, τις τεχνικές και τις μεθοδολογίες ανάλυσής του, στο πλαίσιο του ερευνητικού προγράμματος CGAP μπορεί να αποτελέσει έναυσμα και ενδεχομένως κίνητρο για αυτομόρφωσή τους - διαρκές ζητούμενο της εκπαίδευτικής διαδικασίας – σε τομείς της Βιολογίας που έχουν εφαρμογή στην Ιατρική και όχι μόνο.

Η μελέτη και η ανάλυση του Προγράμματος του Γονιδιώματος του Καρκίνου αποτελεί ένα πρώτο βήμα. Προτείνεται μία εστιασμένη διερεύνηση της εξέλιξης της έρευνας του γονιδιώματος με την εφαρμογή των επιτευγμάτων του Προγράμματος. Επιπρόσθετα, η ταχεία εξέλιξη της τεχνολογίας στην ανάλυση αφενός του γονιδιώματος και του μεταγραφώματος του καρκίνου και αφετέρου των επιγενετικών μεταβολών, έδωσε ώθηση στον σχεδιασμό και υλοποίηση νέων Προγραμμάτων έρευνας στον καρκίνο. Η μελέτη των Προγραμμάτων αυτών και του συνόλου των δεδομένων τους θα μπορούσε να αποτελέσει μια άλλη προοπτική βιβλιογραφικής διερεύνησης στη Μοριακή Βιολογία του καρκίνου.

## BIBΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ

Aldaz M. Serial Analysis of Gene Expression (SAGE) in Cancer Research, Expression Profiling of Human Tumors, 2003; 47-60.  
Baak J.P., Janssen E.A., et al. Genomic and Proteomics-the way forward, Annals of Oncology, 2005;16: ii30-ii44.  
Boon K., Osório E.C., et al. An anatomy of normal and malignant gene expression Proc Natl Acad Sci U S A., 2002; 99(17): 11287-11292.  
Gundem G., Van Loo P., et al. The evolutionary history of lethal metastatic prostate cancer. Nature. 2015;520(7547):353-7.  
Krizman D.B., Wagner L, et al. The Cancer Genome Anatomy Project: EST Sequencing and the Genetics of Cancer Progression, Neoplasia, 1999; 1(2) 101-106.

Lal A., Lash A.E., et al. Advances in Brief, A Public Database for Gene Expression in Human Cancers, Cancer Research, 1999; 59:5403-5407  
Lander S., Lauren M., et al. Initial sequencing and analysis of the human genome, Nature, 2001;409:860-921.  
Lash A.E., Tolstoshev C.M., et al. SAGEmap: A Public Gene Expression Resource, 2000, Cold Spring Harbor Laboratory Press: www.genome.org  
Packer B.R., Yeager M., et al., SNP500Cancer: a public resource for sequence validation and assay development for genetic variation in candidate genes, Nucleic Acids Research, 2006;34:617-621  
Packer B.R., Yeager M., et al., SNP500Cancer: a public resource for sequence validation and assay development for genetic variation in candidate genes, Nucleic Acids Res., 2004; 32: 528-532.  
Patel C., Leethanakul J.S. et al. New Approaches to the Understanding of the Molecular Basis of Oral Cancer, Crit. Rev. Oral Biol. Med., 2001; 12(1):55-63.  
Porter D. A., Krop I.E., et al. A SAGE (Serial Analysis of Gene Expression) View of Breast Tumor Progression, Cancer Res., 2001; 61:5697.  
Riggins G.J., Strausberg R.L., et al. Genome and genetic resources from the Cancer Genome Anatomy Project, Hum Mol Genet., 2001;10(7):663-7.  
Rinn J.L. and Chang H.Y. Genome regulation by long noncoding RNAs. Annu Rev Biochem. 2012;81:145-66.  
Strausberg R.L., Dahl C.A., et al. New Opportunities for Uncovering the Molecular Basis of Cancer, Nat Genet, 1997; 16:415-516.  
Strausberg R.L., Buetow K.H., et al. The Cancer Genome Anatomy Project: Building an annotated gene index, TIG, 2000; 16.  
Strausberg R.L., Greenhut S.F., et al., In silico analysis of cancer through the Cancer Genome Anatomy, Project TRENDSin Cell Biology, 2001;11(11).  
Strausberg R.L. and Riggins G.J. Navigating the human transcriptome, PNAS, 2001; 21: 11837-11838.  
Strausberg R.L., The Cancer Genome Anatomy Project: newresources for reading the molecular signatures of cancer, Journal of Pathology, 2001; 195: 31-40.  
Strausberg R L, Simpson A. J. G., et al. Whole-genome cancer analysis as an approach to deeper understanding of tumour biology . Br J Cancer. 2010 ; 102(2): 243-248.  
Vasmataz G., Essand M., et al. Discovery of three genes specifically expressed in human prostate by expressed sequence tag database analysis, Proc Natl Acad Sci USA,1998; 95:300-304.  
Zhang L., Qian L., et al. Similarity-balanced discriminant neighbor embedding and its application to cancer classification based on gene expression data. Comput Biol Med. 2015;64:236-45.